McKinsey&Company

MCKINSEY GLOBAL INSTITUTE

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

DECEMBER 2016

IN COLLABORATION WITH
MCKINSEY ANALYTICS

HIGHLIGHTS

Organizational challenges

Disruptive business models

Enhanced decision making

MCKINSEY GLOBAL INSTITUTE

CELEBRATING 25 YEARS OF INSIGHT

25

MCKINSEY GLOBAL INSTITUTE

In the 25 years since its founding, the McKinsey Global Institute (MGI) has sought to develop a deeper understanding of the evolving global economy. As the business and economics research arm of McKinsey & Company, MGI aims to provide leaders in the commercial, public, and social sectors with the facts and insights on which to base management and policy decisions. The Lauder Institute at the University of Pennsylvania ranked MGI the world's number-one private-sector think tank in its 2015 Global Think Tank Index.

MGI research combines the disciplines of economics and management, employing the analytical tools of economics with the insights of business leaders. Our "micro-to-macro" methodology examines microeconomic industry trends to better understand the broad macroeconomic forces affecting business strategy and public policy. MGI's in-depth reports have covered more than 20 countries and 30 industries. Current research focuses on six themes: productivity and growth, natural resources, labor markets, the evolution of global financial markets, the economic impact of technology and innovation, and urbanization.

Recent reports have assessed the economic benefits of tackling gender inequality, a new era of global competition, Chinese innovation, and digital globalization. MGI is led by four McKinsey & Company senior partners: Jacques Bughin, James Manyika, Jonathan Woetzel, and Frank Mattern, MGI's chairman. Michael Chui, Susan Lund, Anu Madgavkar, and Jaana Remes serve as MGI partners. Project teams are led by the MGI partners and a group of senior fellows, and include consultants from McKinsey offices around the world. These teams draw on McKinsey's global network of partners and industry and management experts. Input is provided by the MGI Council, which co-leads projects and provides guidance; members are Andres Cadena, Richard Dobbs, Katy George, Rajat Gupta, Eric Hazan, Eric Labaye, Acha Leke, Scott Nyquist, Gary Pinkus, Shirish Sankhe, Oliver Tonby, and Eckart Windhagen. In addition, leading economists, including Nobel laureates, act as research advisers.

The partners of McKinsey fund MGI's research; it is not commissioned by any business, government, or other institution. For further information about MGI and to download reports, please visit **www.mckinsey.com/mgi**.

MCKINSEY ANALYTICS

McKinsey Analytics helps clients achieve better performance through data, working together with them to build analytics-driven organizations and providing end-to-end support covering strategy, operations, data science, implementation, and change management. Engagements range from use-case specific applications to full-scale analytics transformations. Teams of McKinsey consultants, data scientists, and engineers work with clients to identify opportunities, assess available data, define solutions, establish optimal hosting environments, ingest data, develop cutting-edge algorithms, visualize outputs, and assess impact while building capabilities to sustain and expand it.

Copyright © McKinsey & Company 2016

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

DECEMBER 2016

Nicolaus Henke | London
Jacques Bughin | Brussels
Michael Chui | San Francisco
James Manyika | San Francisco
Tamim Saleh | London
Bill Wiseman | Taipei
Guru Sethupathy | Washington, DC

PREFACE

Five years ago, the McKinsey Global Institute (MGI) released *Big data: The next frontier for innovation, competition, and productivity*. In the years since, data science has continued to make rapid advances, particularly on the frontiers of machine learning and deep learning. Organizations now have troves of raw data combined with powerful and sophisticated analytics tools to gain insights that can improve operational performance and create new market opportunities. Most profoundly, their decisions no longer have to be made in the dark or based on gut instinct; they can be based on evidence, experiments, and more accurate forecasts.

As we take stock of the progress that has been made over the past five years, we see that companies are placing big bets on data and analytics. But adapting to an era of more data-driven decision making has not always proven to be a simple proposition for people or organizations. Many are struggling to develop talent, business processes, and organizational muscle to capture real value from analytics. This is becoming a matter of urgency, since analytics prowess is increasingly the basis of industry competition, and the leaders are staking out large advantages. Meanwhile, the technology itself is taking major leaps forward—and the next generation of technologies promises to be even more disruptive. Machine learning and deep learning capabilities have an enormous variety of applications that stretch deep into sectors of the economy that have largely stayed on the sidelines thus far.

This research is a collaboration between MGI and McKinsey Analytics, building on more than five years of research on data and analytics as well as knowledge developed in work with clients across industries. This research also draws on a large body of MGI research on digital technology and its effects on productivity, growth, and competition. It aims to help organizational leaders understand the potential impact of data and analytics, providing greater clarity on what the technology can do and the opportunities at stake.

The research was led by Nicolaus Henke, global leader of McKinsey Analytics, based in London; Jacques Bughin, an MGI director based in Brussels; Michael Chui, an MGI partner based in San Francisco; James Manyika, an MGI director based in San Francisco; Tamim Saleh, a senior partner of McKinsey based in London; and Bill Wiseman, a senior partner of McKinsey based in Taipei. The project team, led by Guru Sethupathy and Andrey Mironenko, included Ville-Pekka Backlund, Rachel Forman, Pete Mulligan, Delwin Olivan, Dennis Schwedhelm, and Cory Turner. Lisa Renaud served as senior editor. Sincere thanks go to our colleagues in operations, production, and external relations, including Tim Beacom, Marisa Carder, Matt Cooke, Deadra Henderson, Richard Johnson, Julie Philpot, Laura Proudlock, Rebeca Robboy, Stacey Schulte, Margo Shimasaki, and Patrick White.

We are grateful to the McKinsey Analytics leaders who provided guidance across the research, including Dilip Bhattacharjee, Alejandro Diaz, Mikael Hagstroem, and Chris Wigley. In addition, this project benefited immensely from the many McKinsey colleagues who shared their expertise and insights. Thanks go to Ali Arat, Matt Ariker, Steven Aronowitz, Bill Aull, Sven Beiker, Michele Bertoncello, James Biggin-Lamming, Yves Boussemart, Chad Bright, Chiara Brocchi, Bede Broome, Alex Brotschi, David Bueno, Eric Buesing, Rune Bundgaard, Sarah Calkins, Ben Cheatham, Joy Chen, Sastry Chilukuri,

Brian Crandall, Zak Cutler, Seth Dalton, Severin Dennhardt, Alexander DiLeonardo, Nicholas Donoghoe, Jonathan Dunn, Leeland Ekstrom, Mehdi El Ouali, Philipp Espel, Matthias Evers, Robert Feldmann, David Frankel, Luke Gerdes, Greg Gilbert, Taras Gorishnyy, Josh Gottlieb, Davide Grande, Daina Graybosch, Ferry Grijpink, Wolfgang Günthner, Vineet Gupta, Markus Hammer, Ludwig Hausmann, Andras Havas, Malte Hippe, Minha Hwang, Alain Imbert, Mirjana Jozic, Hussein Kalaoui, Matthias Kässer, Joshua Katz, Sunil Kishore, Bjorn Kortner, Adi Kumar, Tom Latkovic, Daniel Läubli, Jordan Levine, Nimal Manuel, J.R. Maxwell, Tim McGuire, Doug McElhaney, Fareed Melhem, Phillipe Menu, Brian Milch, Channie Mize, Timo Möller, Stefan Nagel, Deepali Narula, Derek Neilson, Florian Neuhaus, Dimitri Obolenski, Ivan Ostojic, Miklos Radnai, Santiago Restrepo, Farhad Riahi, Stefan Rickert, Emir Roach, Matthias Roggendorf, Marcus Roth, Tom Ruby, Alexandru Rus, Pasha Sarraf, Whitney Schumacher, Jeongmin Seong, Sha Sha, Abdul Wahab Shaikh, Tatiana Sivaeva, Michael Steinmann, Kunal Tanwar, Mike Thompson, Rob Turtle, Jonathan Usuka, Vijay Vaidya, Sri Velamoor, Richard Ward, Khilony Westphely, Dan Williams, Simon Williams, Eckart Windhagen, Martin Wrulich, Ziv Yaar, and Gordon Yu.

Our academic adviser was Martin Baily, Senior Fellow and Bernard L. Schwartz Chair in Economic Policy Development at the Brookings Institution, who challenged our thinking and provided valuable feedback and guidance. We also thank Steve Langdon and the Google TensorFlow group for their helpful feedback on machine learning.

This report contributes to MGI's mission to help business and policy leaders understand the forces transforming the global economy and prepare for the next wave of growth. As with all MGI research, this work is independent, reflects our own views, and has not been commissioned by any business, government, or other institution. We welcome your comments on the research at MGI@mckinsey.com.

Jacques Bughin
Director, McKinsey Global Institute
Senior Partner, McKinsey & Company
Brussels

James Manyika
Director, McKinsey Global Institute
Senior Partner, McKinsey & Company
San Francisco

Jonathan Woetzel
Director, McKinsey Global Institute
Senior Partner, McKinsey & Company
Shanghai

December 2016

© Chombosan/Shutterstock

CONTENTS

HIGHLIGHTS

The demand for talent

Radical personalization in health care

Machine learning and the automation of work activities

In Brief
Page vi

Executive summary
Page 1

1. The data and analytics revolution gains momentum
Page 21

2. Opportunities still uncaptured
Page 29

3. Mapping value in data ecosystems
Page 43

4. Models of disruption fueled by data and analytics
Page 55

5. Deep learning: The coming wave
Page 81

Technical appendix
Page 95

Bibliography
Page 121

IN BRIEF

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

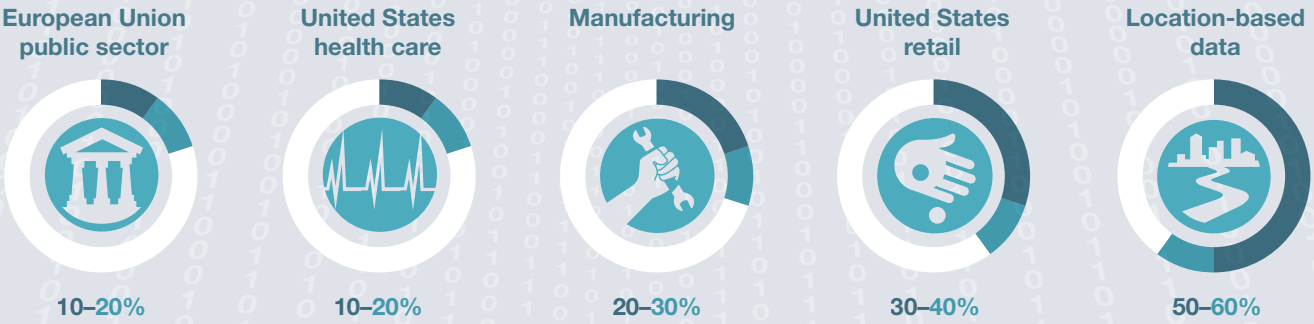
Data and analytics capabilities have made a leap forward in recent years. The volume of available data has grown exponentially, more sophisticated algorithms have been developed, and computational power and storage have steadily improved. The convergence of these trends is fueling rapid technology advances and business disruptions.

- Most companies are capturing only a fraction of the potential value from data and analytics. Our 2011 report estimated this potential in five domains; revisiting them today shows a great deal of value still on the table. The greatest progress has occurred in location-based services and in retail, both areas with digital native competitors. In contrast, manufacturing, the public sector, and health care have captured less than 30 percent of the potential value we highlighted five years ago. Further, new opportunities have arisen since 2011, making the gap between the leaders and laggards even bigger.
- The biggest barriers companies face in extracting value from data and analytics are organizational; many struggle to incorporate data-driven insights into day-to-day business processes. Another challenge is attracting and retaining the right talent—not only data scientists but business translators who combine data savvy with industry and functional expertise.
- Data and analytics are changing the basis of competition. Leading companies are using their capabilities not only to improve their core operations but to launch entirely new business models. The network effects of digital platforms are creating a winner-take-most dynamic in some markets.
- Data is now a critical corporate asset. It comes from the web, billions of phones, sensors, payment systems, cameras, and a huge array of other sources—and its value is tied to its ultimate use. While data itself will become increasingly commoditized, value is likely to accrue to the owners of scarce data, to players that aggregate data in unique ways, and especially to providers of valuable analytics.
- Data and analytics underpin several disruptive models. Introducing new types of data sets ("orthogonal data") can disrupt industries, and massive data integration capabilities can break through organizational and technological silos, enabling new insights and models. Hyperscale digital platforms can match buyers and sellers in real time, transforming inefficient markets. Granular data can be used to personalize products and services—and, most intriguingly, health care. New analytical techniques can fuel discovery and innovation. Above all, data and analytics can enable faster and more evidence-based decision making.
- Recent advances in machine learning can be used to solve a tremendous variety of problems—and deep learning is pushing the boundaries even further. Systems enabled by machine learning can provide customer service, manage logistics, analyze medical records, or even write news stories. The value potential is everywhere, even in industries that have been slow to digitize. These technologies could generate productivity gains and an improved quality of life—along with job losses and other disruptions. Previous MGI research found that 45 percent of work activities could potentially be automated by currently demonstrated technologies; machine learning can be an enabling technology for the automation of 80 percent of those activities. Breakthroughs in natural language processing could expand that impact even further.

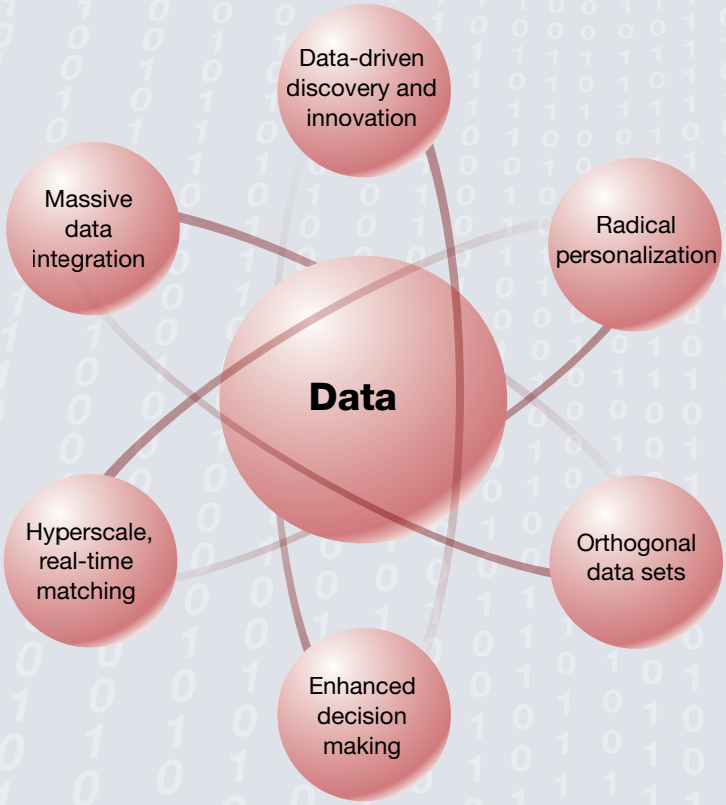
Data and analytics are already shaking up multiple industries, and the effects will only become more pronounced as adoption reaches critical mass. An even bigger wave of change is looming on the horizon as deep learning reaches maturity, giving machines unprecedented capabilities to think, problem-solve, and understand language. Organizations that are able to harness these capabilities effectively will be able to create significant value and differentiate themselves, while others will find themselves increasingly at a disadvantage.

The age of analytics:
Competing in a data-driven world

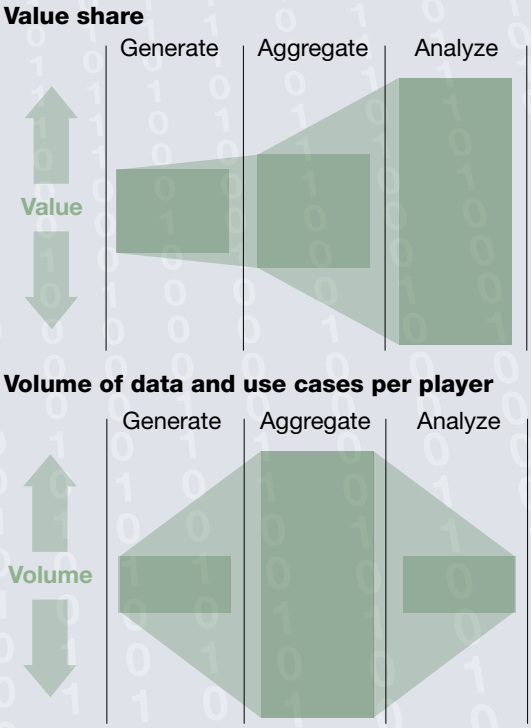
Only a fraction of the value we envisioned in 2011 has been captured to date



Data and analytics fuel 6 disruptive models that change the nature of competition

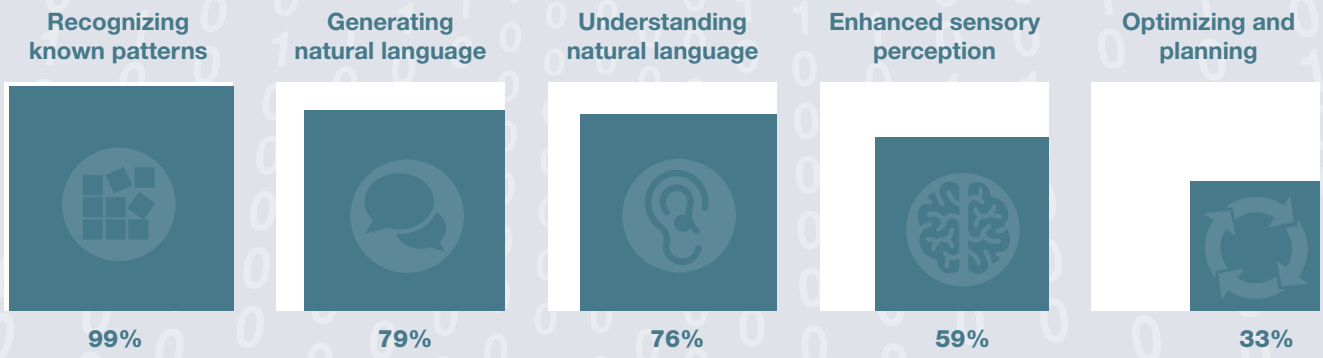


As data ecosystems evolve, value will accrue to providers of analytics, but some data generators and aggregators will have unique value



Machine learning has broad applicability in many common work activities

Percent of work activities that require:



MCKINSEY GLOBAL INSTITUTE

McKinsey&Company

© B. Busco/Getty Images

EXECUTIVE SUMMARY

Back in 2011, the McKinsey Global Institute published a report highlighting the transformational potential of big data.[1] Five years later, we remain convinced that this potential has not been overhyped. In fact, we now believe that our 2011 analyses gave only a partial view. The range of applications and opportunities has grown even larger today.

The convergence of several technology trends is accelerating progress. The volume of data continues to double every three years as information pours in from digital platforms, wireless sensors, and billions of mobile phones. Data storage capacity has increased, while its cost has plummeted. Data scientists now have unprecedented computing power at their disposal, and they are devising ever more sophisticated algorithms.

The companies at the forefront of these trends are using their capabilities to tackle business problems with a whole new mindset. In some cases, they have introduced data-driven business models that have taken entire industries by surprise. Digital natives have an enormous advantage, and to keep up with them, incumbents need to apply data and analytics to the fundamentals of their existing business while simultaneously shifting the basis of competition. In an environment of increasing volatility, legacy organizations need to have one eye on high-risk, high-reward moves of their own, whether that means entering new markets or changing their business models. At the same time, they have to apply analytics to improve their core operations. This may involve identifying new opportunities on the revenue side, using analytics insights to streamline internal processes, and building mechanisms for experimentation to enable continuous learning and feedback.

Organizations that pursue this two-part strategy will be ready to take advantage of opportunities and thwart potential disruptors—and they have to assume that those disruptors are right around the corner. Data and analytics have altered the dynamics in many industries, and change will only accelerate as machine learning and deep learning develop capabilities to think, problem-solve, and understand language. The potential uses of these technologies are remarkably broad, even for sectors that have been slow to digitize. As we enter a world of self-driving cars, personalized medicine, and intelligent robots, there will be enormous new opportunities as well as significant risks—not only for individual companies but for society as a whole.

MOST COMPANIES ARE CAPTURING ONLY A FRACTION OF THE POTENTIAL VALUE OF DATA AND ANALYTICS

Turning a world full of data into a data-driven world is an idea that many companies have found difficult to pull off in practice. Our 2011 report estimated the potential for big data and analytics to create value in five specific domains. Revisiting them today shows both uneven progress and a great deal of that value still on the table (Exhibit E1).

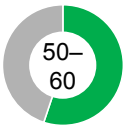
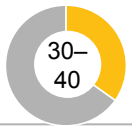
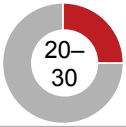
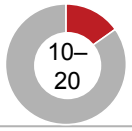
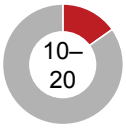
We see the greatest progress in location-based services and in US retail. In contrast, adoption is lagging in manufacturing, the EU public sector, and US health care. Incentive problems and regulatory issues pose additional barriers to adoption in the public sector and health care. In several cases, incumbent stakeholders that would have the most to lose from the kinds of changes data and analytics could enable also have a strong influence on regulations, a factor that could hinder adoption.

---

[1] *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, June 2011.

Exhibit E1

There has been uneven progress in capturing value from data and analytics

	Potential impact: 2011 research	Value captured %	Major barriers
Location-based data	▪ $100 billion+ revenues for service providers ▪ Up to $700 billion value to end users		▪ Penetration of GPS-enabled smartphones globally
US retail[1]	▪ 60%+ increase in net margin ▪ 0.5–1.0% annual productivity growth		▪ Lack of analytical talent ▪ Siloed data within companies
Manufacturing[2]	▪ Up to 50% lower product development cost ▪ Up to 25% lower operating cost ▪ Up to 30% gross margin increase		▪ Siloed data in legacy IT systems ▪ Leadership skeptical of impact
EU public sector[3]	▪ ~€250 billion value per year ▪ ~0.5% annual productivity growth		▪ Lack of analytical talent ▪ Siloed data within different agencies
US health care	▪ $300 billion value per year ▪ ~0.7% annual productivity growth		▪ Need to demonstrate clinical utility to gain acceptance ▪ Interoperability and data sharing

1 Similar observations hold true for the EU retail sector.
2 Manufacturing levers divided by functional application.
3 Similar observations hold true for other high-income country governments.

SOURCE: Expert interviews; McKinsey Global Institute analysis

- **Location-based services:** GPS-enabled smartphones have put mapping technology in the pockets of billions of users. The markets for global positioning system navigation devices and services, mobile phone location-based service applications, and geo-targeted mobile advertising services have reached 50 to 60 percent of the value we envisioned in 2011. End consumers are capturing the lion's share of the benefits, mostly through time and fuel savings as well as new types of mobile services. Beyond the value we envisioned in 2011, there are growing opportunities for businesses to use geospatial data to track assets, teams, and customers across dispersed locations in order to generate new insights and improve efficiency.

- **US retail:** Retailers can mine a trove of transaction-based and behavioral data from their customers. Thin margins (especially in the grocery sector) and pressure from industry-leading early adopters such as Amazon and Walmart have created strong incentives to put that data to work in everything from cross-selling additional products to reducing costs throughout the entire value chain. The US retail sector has realized 30 to 40 percent of the potential margin improvements and productivity growth we envisioned in 2011, but again, a great deal of value has gone to consumers.

- **Manufacturing:** Manufacturing industries have achieved only about 20 to 30 percent of the potential value we estimated in 2011—and most has gone to a handful of industry leaders. Within research and design, design-to-value applications have seen the greatest uptick in adoption, particularly among carmakers. Some industry leaders have developed digital models of the entire production process ("digital factories"). More companies have integrated sensor data-driven operations analytics, often reducing

operating costs by 5 to 15 percent. After-sales servicing offers are beginning to be based on real-time surveillance and predictive maintenance.

10-20%
of the potential value has been captured in the public sector and health care

- **The EU public sector:** Our 2011 report analyzed how the European Union's public sector could use data and analytics to make government services more efficient, reduce fraud and errors in transfer payments, and improve tax collection, potentially achieving some €250 billion worth of annual savings. But only about 10 to 20 percent of this has materialized. Some agencies have moved more interactions online, and many (particularly tax agencies) have introduced pre-filled forms. But across Europe and other advanced economies, adoption and capabilities vary greatly. The complexity of existing systems and the difficulty of attracting scarce analytics talent with public-sector salaries have slowed progress. Despite this, we see even wider potential today for societies to use analytics to make more evidence-based decisions in many aspects of government.

- **US health care:** To date, only 10 to 20 percent of the opportunities we outlined in 2011 have been realized by the US health-care sector. A range of barriers—including a lack of incentives, the difficulty of process and organizational changes, a shortage of technical talent, data-sharing challenges, and regulations—have combined to limit adoption. Within clinical operations, the biggest success has been the shift to electronic medical records, although the vast stores of data they contain have not yet been fully mined. While payers have been slow to capitalize on big data for accounting and pricing, a growing industry now aggregates and synthesizes clinical records, and analytics have taken on new importance in public health surveillance. Many pharmaceutical firms are using analytics in R&D, particularly in streamlining clinical trials. While the health-care sector continues to lag in adoption, there are enormous unrealized opportunities to transform clinical care and deliver personalized medicine (a topic we will return to below).

LEGACY COMPANIES HAVE TO OVERCOME HURDLES TO ACCELERATE THEIR ANALYTICS TRANSFORMATION

The relatively slow pace of progress in some of the domains described above points to the fact that many companies that have begun to deploy data and analytics have not realized the full value. Some have responded to competitive pressure by making large technology investments but have failed to make the organizational changes needed to make the most of them.

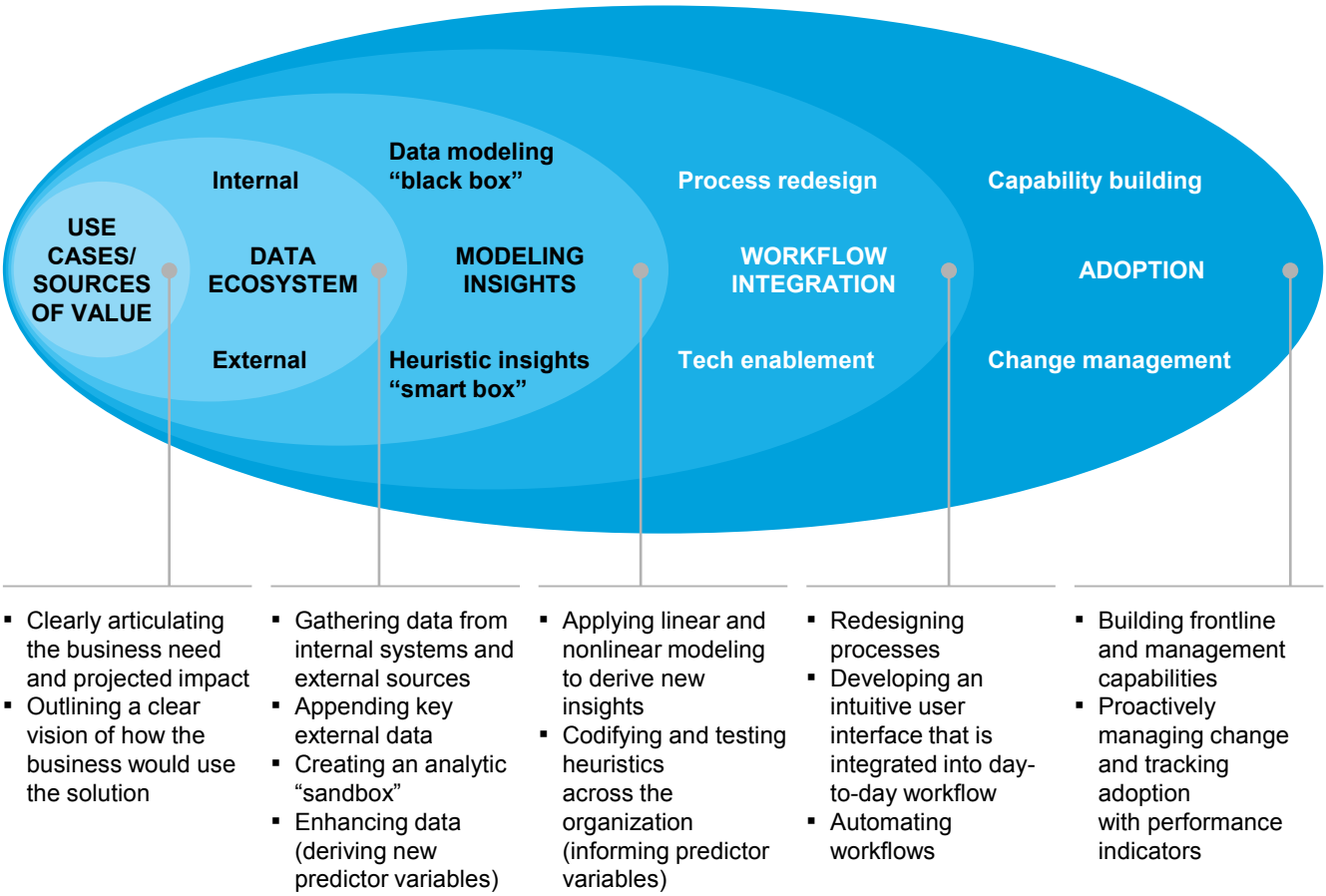
An effective transformation strategy can be broken down into several components (Exhibit E2). The first step should be asking some fundamental questions to shape the strategic vision: What will data and analytics be used for? How will the insights drive value? How will the value be measured? The second element is building out the underlying data architecture as well as data collection or generation capabilities. Many incumbents struggle with switching from legacy data systems to a more nimble and flexible architecture to store and harness big data. They may also need to digitize their operations more fully in order to capture more data from their customer interactions, supply chains, equipment, and internal processes. Looking at a wide variety of indicators that measure digitization, we see a striking gap between leading firms and average firms on this front.[2] The third piece is acquiring the analytics capabilities needed to derive insights from data; organizations may choose to add in-house capabilities or outsource to specialists. The fourth component is a common stumbling block: changing business processes to incorporate data insights into the actual workflow. This requires getting the right data insights into the hands of the right personnel. Finally, organizations need to build the capabilities of executives and mid-level managers to understand how to use data-driven insights—and to begin to rely on them as the basis for making decisions.

[2] *Digital America: A tale of the haves and have-mores*, McKinsey Global Institute, December 2015; and *Digital Europe: Pushing the frontier, capturing the benefits*, McKinsey Global Institute, June 2016.

Where digital native companies were built for analytics, legacy companies have to do the hard work of overhauling or changing existing systems. Neglecting any of these elements can limit the potential value of analytics or even leave an organization vulnerable to being disrupted. It may be a difficult transition, but some long-established names, including GE and Union Pacific, have managed to pull it off.

Exhibit E2

Successful data and analytics transformation requires focusing on five elements


USE CASES/ SOURCES OF VALUE
DATA ECOSYSTEM
Internal
External
MODELING INSIGHTS
Data modeling "black box"
Heuristic insights "smart box"
WORKFLOW INTEGRATION
Process redesign
Tech enablement
ADOPTION
Capability building
Change management
▪ Clearly articulating the business need and projected impact
▪ Outlining a clear vision of how the business would use the solution
▪ Gathering data from internal systems and external sources
▪ Appending key external data
▪ Creating an analytic "sandbox"
▪ Enhancing data (deriving new predictor variables)
▪ Applying linear and nonlinear modeling to derive new insights
▪ Codifying and testing heuristics across the organization (informing predictor variables)
▪ Redesigning processes
▪ Developing an intuitive user interface that is integrated into day-to-day workflow
▪ Automating workflows
▪ Building frontline and management capabilities
▪ Proactively managing change and tracking adoption with performance indicators

SOURCE: McKinsey Analytics; McKinsey Global Institute analysis

THERE IS A CONTINUING SHORTAGE OF ANALYTICS TALENT

Across the board, companies report that finding the right talent is the biggest hurdle they face in trying to integrate data and analytics into their existing operations. In a recent McKinsey & Company survey, approximately half of executives across geographies and industries reported greater difficulty recruiting analytical talent than filling any other kind of role. Forty percent say retention is also an issue.[3]

Data scientists, in particular, are in high demand. Our 2011 report hypothesized that demand for data scientists would outstrip supply. This is in fact what we see in the labor market today, despite the fact that universities are adding data and analytics programs and that other types of training programs are proliferating. Average wages for data scientists in the United States rose by approximately 16 percent a year from 2012 to 2014.[4] This far

[3] "The need to lead in data and analytics," McKinsey & Company survey, McKinsey.com, April 2016, available at http://www.mckinsey.com/business-functions/business-technology/our-insights/the-need-to-lead-in-data-and-analytics. The online survey, conducted in September 2015, garnered responses from more than 500 executives across a variety of regions, industries, and company sizes.

[4] *Beyond the talent shortage: How tech candidates search for jobs*, Indeed.com, September 2015.

outstrips the less than 2 percent increase in nominal average wages across all occupations in US Bureau of Labor Statistics data. The scarcity of elite data scientists has even been a factor in some acquisitions of cutting-edge artificial intelligence (AI) startups; deals can command around $5 million to $10 million per employee.

This trend is likely to continue in the near term. While we estimate that the number of graduates from data science programs could increase by a robust 7 percent per year, our high-case scenario projects even greater (12 percent) annual growth in demand, which would lead to a shortfall of some 250,000 data scientists. But a countervailing force could ease this imbalance in the medium term: data preparation, which accounts for more than 50 percent of data science work, could be automated. Whether that dampens the demand for data scientists or simply enables data scientists to shift their work toward analysis and other activities remains to be seen.

Many organizations focus on the need for data scientists, assuming their presence alone will enable an analytics transformation. But another equally vital role is that of the business translator who serves as the link between analytical talent and practical applications to business questions. In addition to being data savvy, business translators need to have deep organizational knowledge and industry or functional expertise. This enables them to ask the data science team the right questions and to derive the right insights from their findings. It may be possible to outsource analytics activities, but business translator roles require proprietary knowledge and should be more deeply embedded into the organization. Many organizations are building these capabilities from within.

2M–4M
projected US demand for business translators over the next decade

We estimate there could be demand for approximately two million to four million business translators in the United States alone over the next decade. Given the roughly 9.5 million US graduates in business and in the STEM fields of science, technology, engineering, and mathematics expected over the same period, nearly 20 to 40 percent of these graduates would need to go into business translator roles to meet demand.[5] Today that figure is only about 10 percent. To reduce this mismatch, wages may have to increase, or more companies will need to implement their own training programs.[6]

As data grows more complex, distilling it and bringing it to life through visualization is becoming critical to help make the results of data analyses digestible for decision makers. We estimate that demand for visualization grew roughly 50 percent annually from 2010 to 2015.[7] In many instances today, organizations are seeking data scientist or business translator candidates who can also execute visualizations. However, we expect that medium-size and large organizations, as well as analytics service providers, will increasingly create specialized positions for candidates who combine a strong understanding of data with user interface, user experience, and graphic design skills.

ANALYTICS LEADERS ARE CHANGING THE NATURE OF COMPETITION AND CONSOLIDATING BIG ADVANTAGES

There are now major disparities in performance between a small group of technology leaders and the average company—in some cases creating winner-take-most dynamics. Leaders such as Apple, Alphabet/Google, Amazon, Facebook, Microsoft, GE, and Alibaba Group have established themselves as some of the most valuable companies in the world.[8] The same trend can be seen among privately held companies. The leading global “unicorns”

5 Non-STEM graduates with quantitative skills can also fill business translator roles.

6 Sam Ransbotham, David Kiron, and Pamela Kirk Prentice, “The talent dividend: Analytics talent is driving competitive advantage at data-oriented companies,” *MIT Sloan Management Review*, April 25, 2015.

7 Based on using the Burning Glass job postings database to search for postings including any of the following skills: data visualization, Tableau, Qlikview, and Spotfire. Normalized with the total number of job postings.

8 Michael Chui and James Manyika, “Competition at the digital edge: ‘Hyperscale businesses,’” *McKinsey Quarterly*, March 2015.

tend to be companies with business models predicated on data and analytics, such as Uber, Lyft, Didi Chuxing, Palantir, Flipkart, Airbnb, DJI, Snapchat, Pinterest, BlaBlaCar, and Spotify. These companies differentiate themselves through their data and analytics assets, processes, and strategies.

The relative value of various assets has shifted. Where previous titans of industry poured billions into factories and equipment, the new leaders invest heavily in digital platforms, data, and analytical talent. New digital native players can circumvent traditional barriers to entry, such as the need to build traditional fixed assets, which enables them to enter markets with surprising speed. Amazon challenged the rest of the retail sector without building stores (though it does have a highly digitized physical distribution network), "fintechs" are providing financial services without physical bank branches, Netflix is changing the media landscape without connecting cables to customers' homes, and Airbnb has introduced a radical new model in the hospitality sector without building hotels. But some digital natives are now erecting new barriers to entry themselves; platforms may have such strong network effects that they give operators a formidable advantage within a given market.

The leading firms have a remarkable depth of analytical talent deployed on a variety of problems—and they are actively looking for ways to enter other industries. These companies can take advantage of their scale and data insights to add new business lines, and those expansions are increasingly blurring traditional sector boundaries.[9] Apple and Alibaba, for instance, have introduced financial products and services, while Google is developing autonomous cars. The importance of data has also upended the traditional relationship between organizations and their customers since every interaction generates information. Sometimes the data itself is so prized that companies offer free services in order to obtain it; this is the case with Facebook, LinkedIn, Pinterest, Twitter, Tencent, and many others. An underlying barter system is at work, particularly in the consumer space, as individuals gain access to digital services in return for data about their behaviors and transactions.

THE VALUE OF DATA DEPENDS ON ITS ULTIMATE USE, AND ECOSYSTEMS ARE EVOLVING TO HELP COMPANIES CAPTURE THAT VALUE

Data is at the heart of the disruptions occurring across the economy. It has become a critical corporate asset, and business leaders want to know what the information they hold is worth. But its value is tied to how it will be used and by whom. A piece of data may yield nothing, or it may yield the key to launching a new product line or cracking a scientific question. It might affect only a small percentage of a company's revenue today, but it could be a driver of growth in the future.

Not all data are created equal

Part of the challenge in valuing data is its sheer diversity. Some of the broad categories include behavioral data (capturing actions in both digital and physical environments), transactional data (records of business dealings), ambient or environmental data (conditions in the physical world monitored and captured by sensors), geospatial data, reference material or knowledge (news stories, textbooks, reference works, literature, and the like), and public records. Some data are structured (that it, easily expressed in rows and columns), while images, audio, and video are unstructured. Data can also come from the web, social media, industrial sensors, payment systems, cameras, wearable devices, and human entry. Billions of mobile phones, in particular, are capturing images, video, and location data. On the demand side, data can provide insights for diverse uses, some of which are more valuable than others.

[9] *Playing to win: The new global competition for corporate profits*, McKinsey Global Institute, September 2015.

Over the long term, value will likely accrue to providers of analytics and data platform owners

Many organizations are hungry to use data to grow and improve performance—and multiple players see market opportunities in this explosion of demand. There are typically many steps between raw data and actual usage, and there are openings to add value at various points along the way. To simplify, we focused on three categories of players in the data ecosystem, recognizing that some players might fill more than one role.

- **Data generation and collection:** The source and platform where data are initially captured.
- **Data aggregation:** Processes and platforms for combining data from multiple sources.
- **Data analysis:** The gleaning of insights from data that can be acted upon.

Usually, the biggest opportunities are unlikely to be in directly monetizing data. As data become easier to collect and as storage costs go down, most data are becoming more commoditized. Proxies now exist for data that were once scarce; Google Trends, for instance, offers a free proxy for public sentiment data that previously would have been collected through phone surveys.

However, there are important exceptions to the commoditization trend. When access is limited by physical barriers or collection is expensive, data will hold its value. An important case in which value can accrue to data generation and collection involves market-making or social media platforms with strong network effects. In certain arenas, a small number of players establish such critical mass that they are in a position to collect and own the vast majority of user behavior data generated in these ecosystems. But in the absence of these types of exceptional supply constraints, simply selling raw data is likely to generate diminishing returns over time.

Another role in the data ecosystem involves aggregating information from different sources. In general, this capability is becoming more accessible and less expensive, but this role can be valuable when certain conditions apply. Data aggregation adds value when combining, processing, and aggregating data is technically difficult or organizationally challenging (for example, when aggregating involves coordinating access across diverse sources). Some companies have built business models around serving as third-party aggregators for competitors within a given industry, and this model has the potential to create network effects as well.

The third part of the data ecosystem, analytics, is where we expect to see the biggest opportunities in the future. The provider of analytics understands the value being generated by those insights and is thus best positioned to capture a portion of that value. Data analytics tools, like other software, already command large margins. Combining analytical tools with business insights for decision makers is likely to multiply the value even further. Increasingly complex data and analytics will require sophisticated translation, and use cases will be very firm-specific. Bad analysis can destroy the potential value of high-quality data, while great analysis can squeeze insights from even mediocre data. In addition, the scarcity of analytics talent is driving up the cost of these services. Given the size of the opportunities, firms in other parts of the ecosystem are scrambling to stake out a niche in the analytics market. Data aggregators are offering to integrate clients' data and perform analysis as a service. One-stop shops offering integrated technology stacks are adding analytics capabilities, such as IBM Watson, as are other professional services and business intelligence firms.

SIX DISRUPTIVE DATA-DRIVEN MODELS AND CAPABILITIES ARE RESHAPING SOME INDUSTRIES—AND COULD SOON TRANSFORM MANY MORE

Certain characteristics of a given market (such as inefficient matching, information asymmetries, and human biases and errors) open the door to disruption. They set the stage for six archetypes to have a major effect (Exhibit E3). In each of these models, the introduction of new data is a key enabler.

Exhibit E3

Data and analytics underpin six disruptive models, and certain characteristics make individual domains susceptible

Indicators of potential for disruption:

- Assets are underutilized due to inefficient signaling
- Supply/demand mismatch
- Dependence on large amounts of personalized data
- Data is siloed or fragmented
- Large value in combining data from multiple sources
- R&D is core to the business model
- Decision making is subject to human biases
- Speed of decision making limited by human constraints
- Large value associated with improving accuracy of prediction

Archetype of disruption	Domains that could be disrupted
Business models enabled by orthogonal data	▪ Insurance ▪ Health care ▪ Human capital/talent
Hyperscale, real-time matching	▪ Transportation and logistics ▪ Automotive ▪ Smart cities and infrastructure
Radical personalization	▪ Health care ▪ Retail ▪ Media ▪ Education
Massive data integration capabilities	▪ Banking ▪ Insurance ▪ Public sector ▪ Human capital/talent
Data-driven discovery	▪ Life sciences and pharmaceuticals ▪ Material sciences ▪ Technology
Enhanced decision making	▪ Smart cities ▪ Health care ▪ Insurance ▪ Human capital/talent

SOURCE: McKinsey Global Institute analysis

Bringing in orthogonal data can change the basis of competition

As data proliferate, many new types, from new sources, can be brought to bear on any problem. In industries where most incumbents have become used to relying on a certain kind of standardized data to make decisions, bringing in fresh types of data sets to supplement those already in use can change the basis of competition. New entrants with privileged access to these "orthogonal" data sets can pose a uniquely powerful challenge to incumbents. We see this playing out in property and casualty insurance, where new companies have entered the marketplace with telematics data that provides insight into driving behavior. This is orthogonal to the demographic data that had previously been used for underwriting. Other domains could be fertile ground for bringing in orthogonal data from the internet of things (IoT). Connected light fixtures, which sense the presence of people in a room and have been sold with the promise of reducing energy usage, generate "data exhaust" that property managers can use to optimize physical space planning. Even in human resources, some organizations have secured employee buy-in to wear devices that capture data and yield insights into the "real" social networks that exist in the workplace, enabling these organizations to optimize collaboration through changes in work spaces.

Orthogonal data will rarely replace the data that are already in use in a domain; it is more likely that an organization will integrate orthogonal data with existing data. Within the other

archetypes below are several examples of orthogonal data being combined with existing data to create new business models and improve performance.

Hyperscale platforms can match supply and demand in real time

Digital platforms provide marketplaces that connect sellers and buyers for many products and services. Some platform operators using data and analytics to do this in real time and on an unprecedented scale—and this can be transformative in markets where supply and demand matching has been inefficient.

In personal transportation, ride-sharing services use geospatial mapping technology to collect crucial data about the precise location of passengers and available drivers in real time. The introduction of this new type of data enabled efficient and instant matching, a crucial innovation in this market. In addition, the data can be analyzed at the aggregate level for dynamic pricing adjustments to help supply and demand adjust. The typical personally owned car is estimated to sit idle 85 to 95 percent of the time, making it a hugely underutilized asset. Platforms such as Uber, Lyft, and Chinese ride-sharing giant Didi Chuxing have been able to expand rapidly without acquiring huge fleets themselves, making it easy for new drivers to put their own underutilized assets to work.

Up to
$2.5T
potential economic impact from continued adoption of mobility services by 2025

By 2030 mobility services, such as ride sharing and car sharing, could account for more than 15 to 20 percent of total passenger vehicle miles globally. This growth—and the resulting hit to the taxi industry—may be only a hint of what is to come. Automakers are the biggest question mark. While sales will likely continue to grow in absolute numbers, we estimate that the shift toward mobility services could halve the growth rate of global vehicle sales by 2030. Consumers could save on car purchases, fuel, and parking. If mobility services attain 10 to 30 percent adoption among low-mileage urban vehicle users, the ensuing economic impact could reach $845 billion to some $2.5 trillion globally by 2025. Some of this value will surely go to consumer surplus, while some will go to the providers of these platforms and mobility services.

Data and analytics enable “radical personalization”

Data and analytics can reveal finer levels of distinctions, and one of the most powerful uses is micro-segmenting a population based on the characteristics of individuals. Using the resulting insights to personalize products and services on a wide scale is changing the fundamentals of competition in many sectors, including education, travel and leisure, media, retail, and advertising.

This capability could have profound implications for the way health care is delivered if the sector can incorporate the behavioral, genetic, and molecular data connected with many individual patients. The declining costs of genome sequencing, the advent of proteomics, and the growth of real-time monitoring technologies make it possible to generate this kind of new, ultra-granular data. These data can reshape health care in two profound ways. First, they can help address information asymmetries and incentive problems in the health-care system. Now that a more complete view of the patient is available, incentives could be changed for hospitals and other providers to shift their focus from disease treatment to wellness and prevention, saving huge sums on medical expenditures and improving the quality of life. Second, having more granular and complete data on individual patients can make treatments more precise. Pharmaceutical and medical device companies have enormous possibilities in R&D for accelerating drug discovery, although they will be challenged to create new business models to deliver treatments tailored to smaller, targeted patient populations. Treatments, dosages, and care settings can be personalized to individuals, leading to more effective outcomes with fewer side effects and reduced costs.

Personalized medicine could reduce health-care costs while allowing people to enjoy longer, healthier, and more productive lives. The total impact could range from $2 trillion

to $10 trillion. The wide range depends on the many uncertainties involved, including how rapidly the health-care system can adapt and whether R&D applications produce breakthrough treatments.

Massive data integration capabilities can break down organizational silos

The first step in creating value from data and analytics is accessing all the information that is relevant to a given problem. This may involve generating the data, accessing it from new sources, breaking silos within an organization to link existing data, or all of the above. Combining and integrating large stores of data from all of these varied sources has incredible potential to yield insights, but many organizations have struggled with creating the right structure for that synthesis to take place.

Up to
$260B
potential global impact of massive data integration in retail banking

Retail banking, for instance, is an industry rich with data on customers' transactions, financial status, and demographics. But few institutions have made the most of the data due to internal barriers and the variable quality of the information itself. Surmounting these barriers is critical now that social media, call center discussions, video footage from branches, and data acquired from external sources and partners can be used to form a more complete picture of customers. Massive data integration has significant potential for retail banks. It can enable better cross-selling, the development of personalized products, dynamic pricing, better risk assessment, and more effective marketing—and it can help firms achieve more competitive cost structures than many incumbent institutions. All told, we estimate a potential economic impact of $110 billion to $170 billion in the retail banking industry in developed markets and approximately $60 billion to $90 billion in emerging markets.

Additionally, companies in other sectors can become part of the financial services ecosystem if they bring in orthogonal data—such as non-financial data that provides a more comprehensive and detailed view of the customer. These players may have large customer bases and advanced analytics capabilities created for their core businesses, and they can use these advantages to make rapid moves across sector boundaries. Alibaba's creation of Alipay and Apple's unveiling of Apple Pay are prime examples of this trend.

Data and analytics can fuel discovery and innovation

One of the main components of productivity growth, innovation can be applied to both processes and products. Throughout history, innovative ideas have sprung from human ingenuity and creativity—but now data and algorithms can support, enhance, or even replace human ingenuity in some instances.

In the realm of process innovation, data and analytics are helping organizations determine how to structure teams, resources, and workflows. High-performing teams can be many times more productive than low-performing teams, so understanding this variance and how to build more effective collaboration is a huge opportunity for organizations. This involves looking at issues such as the complementarity of skills, optimal team sizes, whether teams need to work together in person, what past experience or training is important, and even how their personalities may mesh. Data and analytics can test hypotheses and find new patterns that may not have even occurred to managers. Vast amounts of email, calendar, locational, and other data are available to understand how people work together and communicate, all of which can lead to new insights about improving performance.

In product innovation, data and analytics can transform research and development in areas such as materials science, synthetic biology, and life sciences. Leading pharmaceutical companies are using data and analytics to aid with drug discovery. Data from a variety of sources could better determine the chemical compounds that would serves as effective drug treatments for a variety of diseases. AstraZeneca and Human Longevity are partnering

to build a database of one million genomic and health records along with 500,000 DNA samples from clinical trials. The associations and patterns that can be gleaned from that data could prove to be immensely valuable in advancing scientific and drug development breakthroughs.

Algorithms can support and enhance human decision making

When humans make decisions, the process is often muddy, biased, or limited by our inability to process information overload. Data and analytics can change all that by bringing in more data points from new sources, breaking down information asymmetries, and adding automated algorithms to make the process instantaneous. As the sources of data grow richer and more diverse, there are many ways to use the resulting insights to make decisions faster, more accurate, more consistent, and more transparent.

There are many examples of how this can play out in industries and domains across the economy. Smart cities, for example, are one of the most promising settings for applying the ability of machines and algorithms to process huge quantities of information in a fraction of the time it takes humans. Using sensors to improve traffic flows and the internet of things to enable utilities to reduce waste and keep infrastructure systems working at top efficiency are just two of the myriad possible municipal applications. One of the most promising applications of data and analytics is in the prevention of medical errors. Advanced analytical support tools can flag potential allergies or dangerous drug interactions for doctors and pharmacists alike, ensuring that their decisions are consistent and reliable. And finally, perhaps no area of human decision making is quite as opaque and clouded by asymmetric information as hiring. Data and analytics have the potential to create a more transparent labor market by giving employers and job seekers access to data on the supply and demand for particular skills, the wages associated with various jobs, and the value of different degree programs.

THE FRONTIERS OF MACHINE LEARNING, INCLUDING DEEP LEARNING, HAVE RELEVANCE IN EVERY INDUSTRY AND WIDE-RANGING POTENTIAL TO SOLVE PROBLEMS

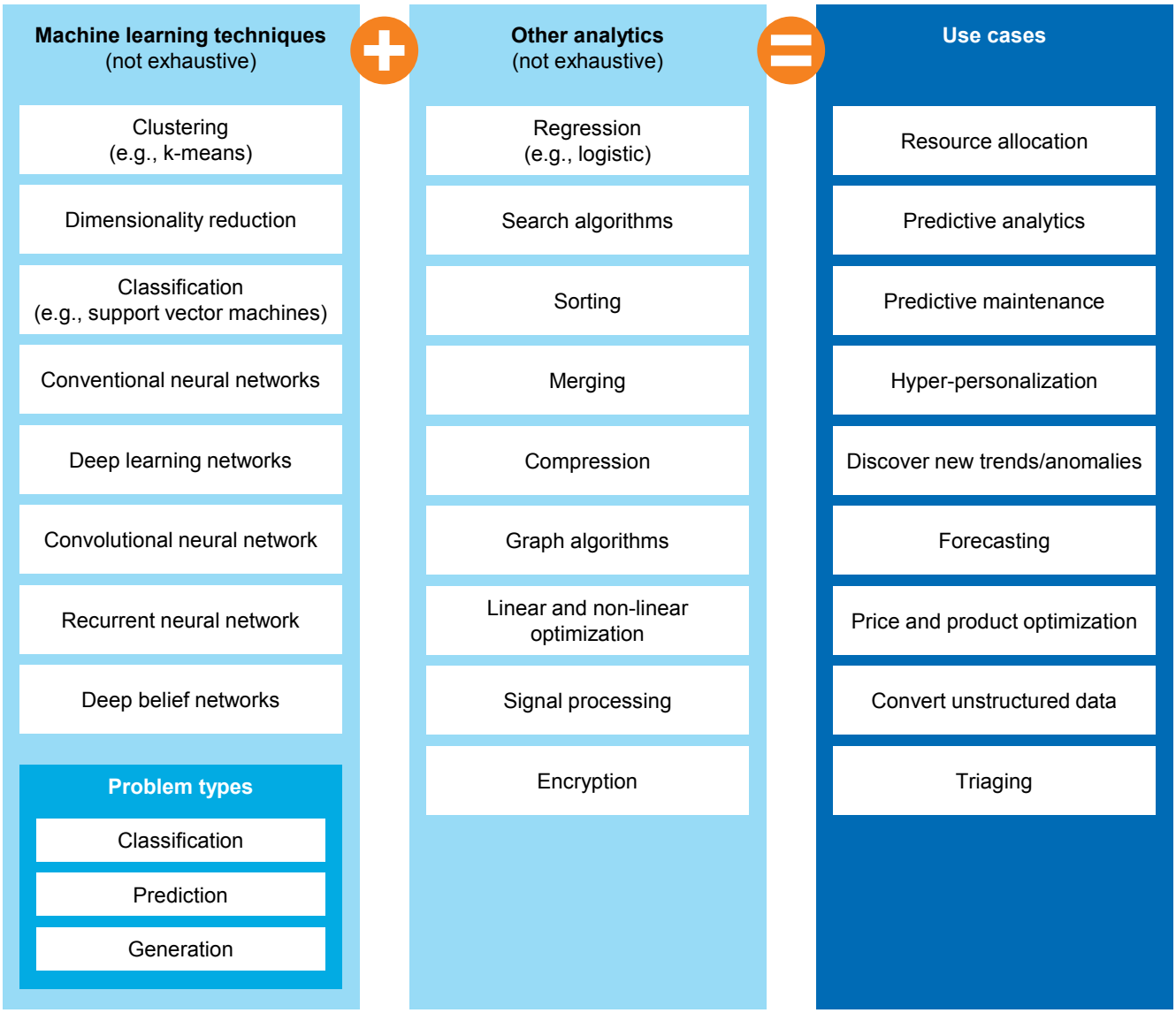
Machine learning can enhance the power of each of the archetypes described above. Conventional software programs are hard-coded by humans with specific instructions on the tasks they need to execute. By contrast, it is possible to create algorithms that "learn" from data without being explicitly programmed. The concept underpinning machine learning is to give the algorithm a massive number of "experiences" (training data) and a generalized strategy for learning, then let it identify patterns, associations, and insights from the data. In short, these systems are trained rather than programmed.

Some machine learning techniques, such as regressions, support vector machines, and k-means clustering, have been in use for decades. Others, while developed previously, have become viable only now that vast quantities of data and unprecedented processing power are available. Deep learning, a frontier area of research within machine learning, uses neural networks with many layers (hence the label "deep") to push the boundaries of machine capabilities. Data scientists have recently made breakthroughs using deep learning to recognize objects and faces and to understand and generate language. Reinforcement learning is used to identify the best actions to take now in order to reach some future goal. These type of problems are common in games but can be useful for solving dynamic optimization and control theory problems—exactly the type of issues that come up in modeling complex systems in fields such as engineering and economics. Transfer learning focuses on storing knowledge gained while solving one problem and applying it to a different problem. Machine learning, combined with other techniques, could have an enormous range of uses (see Exhibit E4 and Box E1, "The impact of machine learning").

This research offers a broad initial exploration of machine learning through two lenses. First, we investigate which business uses across 12 industries could be met by machine learning. Second, we examine which work activities currently performed by people could potentially be automated through machine learning and how that could play out across occupations. The initial findings here are meant to set the stage for future research.

Exhibit E4

Machine learning can be combined with other types of analytics to solve a large swath of business problems



SOURCE: McKinsey Global Institute analysis

Understanding the capabilities of machine learning and deep learning

Machine learning capabilities are best suited for solving three broad categories of problems: classification, prediction/estimation, and generation (Exhibit E5). Classification problems are about observing the world, including identifying objects in images and video, and recognizing text and audio. Classification also involves finding associations in data or segmenting it into clusters, which is useful in tasks such as customer segmentation. Machine learning can also be used to predict the likelihood of events and forecast outcomes. Lastly, it can be used to generate content, from interpolating missing data to generating the next frame in a video sequence.

Exhibit E5

Machine learning can help solve classification, prediction, and generation problems

Classification	Classify/label visual objects	Identify objects, faces in images and video
	Classify/label writing and text	Identify letters, symbols, words in writing sample
	Classify/label audio	Classify and label songs from audio samples
	Cluster, group other data	Segment objects (e.g., customers, product features) into categories, clusters
	Discover associations	Identify that people who watch certain TV shows also read certain books
Prediction	Predict probability of outcomes	Predict the probability that a customer will choose another provider
	Forecast	Trained on historical data, forecast demand for a product
	Value function estimation	Trained on thousands of games played, predict/estimate rewards from actions from future states for dynamic games
Generation	Generate visual objects	Trained on a set of artist's paintings, generate a new painting in the same style
	Generate writing and text	Trained on a historical text, fill in missing parts of a single page
	Generate audio	Generate a new potential recording in the same style/genre
	Generate other data	Trained on certain countries' weather data, fill in missing data points for countries with low data quality

SOURCE: McKinsey Global Institute analysis

Box E1. The impact of machine learning

Machine learning can be applied to a tremendous variety of problems—from keeping race cars running at peak performance to ferreting out fraud.

Off the track, Formula One (F1) teams compete in an arms race to make their cars faster. Top F1 teams pour hundreds of millions of dollars annually into development, continually aiming for incremental technological improvements that can boost speed. With so much at stake, F1 engineering teams constantly seek to improve productivity. Three F1 teams recently turned to machine learning to hold down costs in their aerodynamics operations divisions, which typically eat up more than 80 percent of development resources. Building on years of diverse project data—including CAD logs, human resources data, and employee communications—they looked for patterns that influenced the efficiency of an individual project. They discovered, for example, that too many engineers or long stoppages typically increased labor hours on a given project by 5 to 6 percent, while team use of the documentation system improved productivity by more than 4 percent. Overall, this application reduced the budget by 12 to 18 percent, saving millions of dollars.

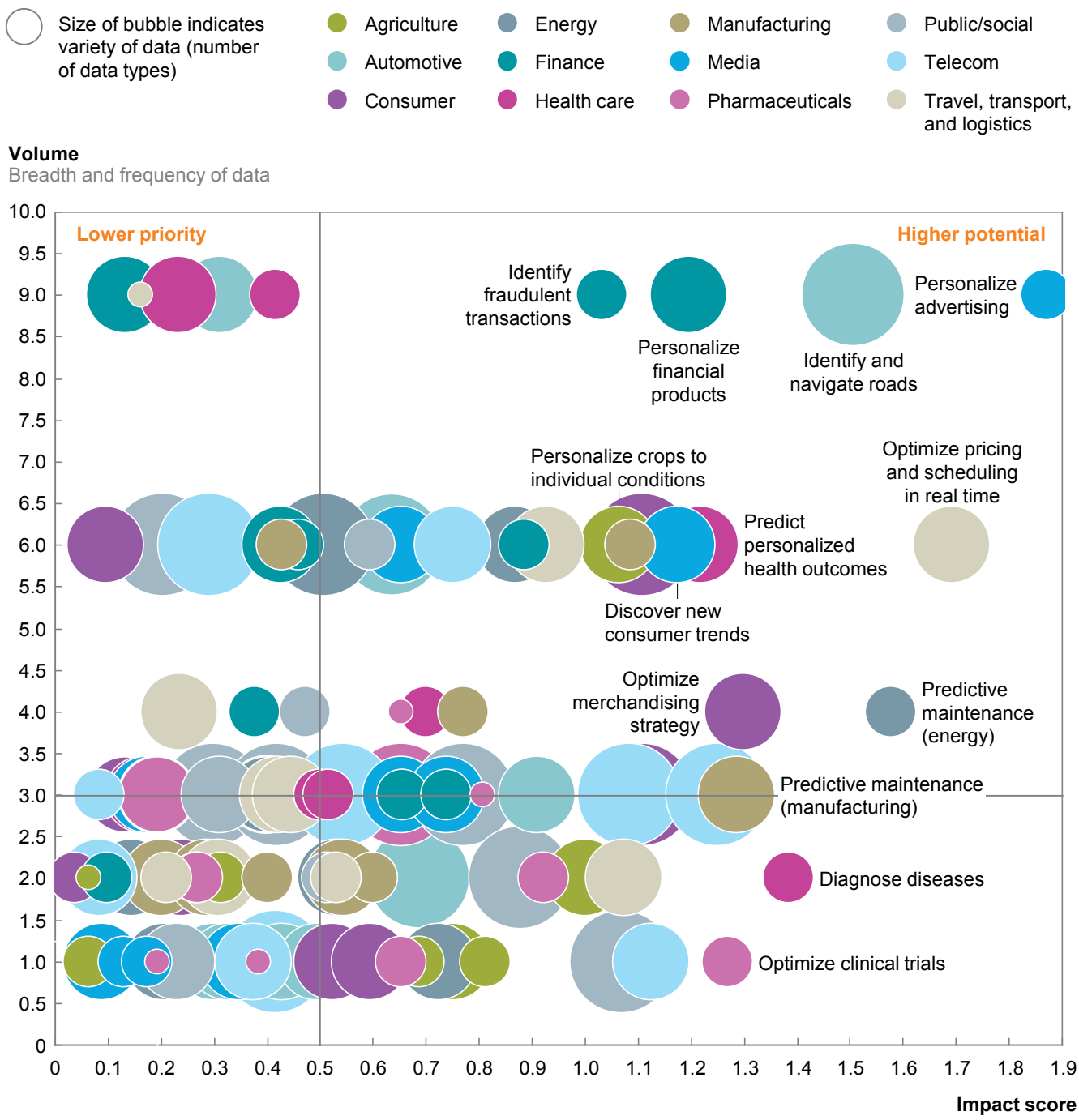
Another application of machine learning, predictive analytics, has proven to be effective at spotting fraud. At one large auto insurer, high accident rates for new client policies suggested that claims were being filed for pre-existing damage. The machine learning model was able to use diverse data to identify groups of new policies with accident rates six times those of the median. This grouping formed the basis of a new pricing strategy that improved profitability by more than 10 percent. Separately, a large retail bank in the United Kingdom used machine learning algorithms to identify fraudulent transactions with more than 90 percent accuracy. In another example, a large payment processor deployed machine learning on its extensive transaction data to identify "mule accounts" involved in money laundering.

We identified 120 potential use cases of machine learning in 12 industries, and surveyed more than 600 industry experts on their potential impact. The most striking finding was the extraordinary breadth of the potential applications of machine learning; each of the use cases was identified as being one of the top three in an industry by at least one expert in that industry. But there were differences.

We plotted the top 120 use cases in Exhibit E6. The y-axis shows the volume of available data (encompassing its breadth and frequency), while the x-axis shows the potential impact, based on surveys of more than 600 industry experts. The size of the bubble reflects the diversity of the available data sources.

Exhibit E6

Machine learning has broad potential across industries and use cases



SOURCE: McKinsey Global Institute analysis

The industry-specific uses that combine data richness with a larger opportunity are the largest bubbles in the top right quadrant of the chart. These represent areas where organizations should prioritize the use of machine learning and prepare for a transformation to take place. Some of the highest-opportunity use cases include personalized advertising; autonomous vehicles; optimizing pricing, routing, and scheduling based on real-time data in travel and logistics; predicting personalized health outcomes; and optimizing merchandising strategy in retail.

The use cases in the top right quadrant fall into four main categories. First is the radical personalization of products and services for customers in sectors such as consumer packaged goods, finance and insurance, health care, and media—an opportunity that most companies have yet to fully exploit. The second is predictive analytics. This includes examples such as triaging customer service calls; segmenting customers based on risk, churn, and purchasing patterns; identifying fraud and anomalies in banking and cybersecurity; and diagnosing diseases from scans, biopsies, and other data. The third category is strategic optimization, which includes uses such as merchandising and shelf optimization in retail, scheduling and assigning frontline workers, and optimizing teams and other resources across geographies and accounts. The fourth category is optimizing operations and logistics in real time, which includes automating plants and machinery to reduce errors and improve efficiency, and optimizing supply chain management.

Advances in deep learning could greatly expand the scope of automation

Previous MGI research examined the potential to automate 2,000 work activities performed in every occupation in the economy.[10] For each work activity, we identified the required level of machine performance across 18 human capabilities that could potentially enable automation.

Machine learning is particularly well-suited to implement seven of those 18 capabilities (Exhibit E7). The first striking observation is that almost all activities require some capabilities that correlate with what machine learning can do. In fact, only four out of more than 2,000 detailed work activities (or 0.2 percent) do not require any of the seven machine learning capabilities. Recognizing known patterns, by itself, is needed in 99 percent of all activities to varying degrees. This is not to say that such a high share of jobs is likely to be automated, but it does underscore the wide applicability of machine learning in many workplaces.

$3T
wages potentially affected if machine learning gains better capabilities in natural language understanding

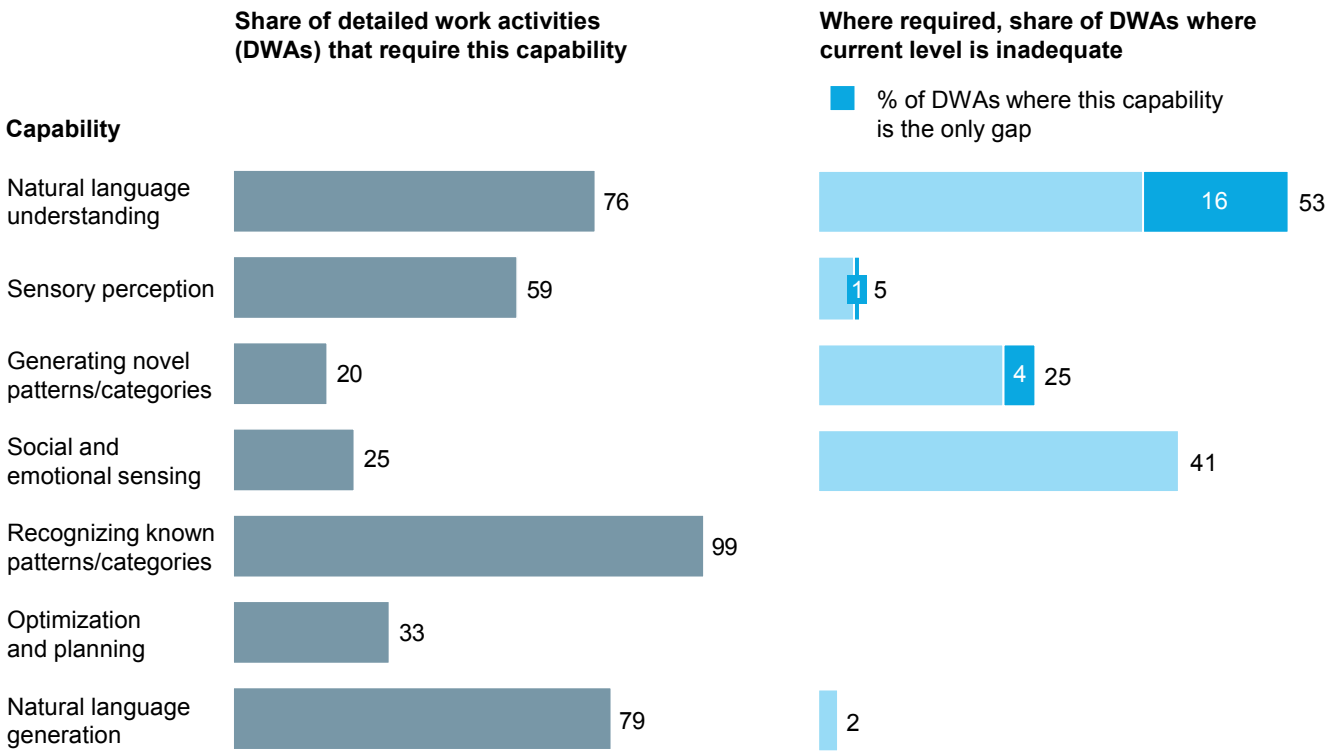
MGI's previous research on automation found that 45 percent of all work activities, associated with $14.6 trillion of wages globally, have the potential to be automated by adapting currently demonstrated technology. Some 80 percent of that could be implemented by using existing machine learning capabilities. But deep learning is in its early stages. Improvements in its capabilities, particularly in natural language understanding, suggest the potential for an even greater degree of automation. In 16 percent of work activities that require the use of language, for example, increasing the performance of machine learning in natural language understanding is the only barrier to automation. Improving natural language capabilities alone could lead to an additional $3 trillion in potential global wage impact.

[10] These "detailed work activities" are defined by O*NET, a data collection program sponsored by the US Department of Labor. See Michael Chui, James Manyika, and Mehdi Miremadi, "Four fundamentals of workplace automation," *McKinsey Quarterly*, November 2015.

Exhibit E7

Improvements in natural learning understanding and generation as well as social sensing would have the biggest impact on expanding the number of work activities that deep learning could technically automate

%



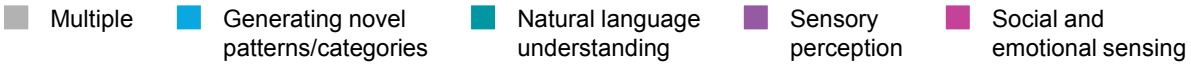
SOURCE: McKinsey Global Institute analysis

We further looked at which occupations that could be affected by improvements in deep learning represent the greatest potential wage impact (Exhibit E8). The role of customer service representatives, in particular, lends itself to automation across most of its work activities. Deep learning is also likely to have a large impact on frontline supervisory roles and in occupations with primarily administrative duties, including executive assistants, cashiers, and waitstaff. Large numbers of people are employed in these occupations, which points to the possibility of substantial job displacement. In addition, advances in machine learning could automate significant percentages of the activities associated with some high-paying jobs such as lawyers and nurses.

While machine learning in general and deep learning in particular have exciting and wide-ranging potential, there are real concerns associated with their development and potential deployment. Some of these, such as privacy, data security, and data ownership, were present even before the big data age. But today new questions have formed.

Exhibit E8

Improvements in deep learning (DL) could affect billions of dollars in wages in ten occupations globally



Occupations	% of time spent on activities that could be automated if DL improves (by DWA group)[1]	Most frequently performed group of DWAs that could be automated if DL improves	Global employ-ment Million	Hourly wage $	Global wages that DL could automate $ billion
Secretaries and administrative assistants, except legal, medical, and executive	28	Interacting with computers to enter data, process information, etc.	48.2	3.90	109.8
Business operations specialists, all other	30	Performing administrative activities	6.1	24.68	94.2
Managers, all other	27	Monitoring processes, materials, or surroundings	8.3	18.25	86.7
First-line supervisors of office and administrative support workers	35	Interpreting the meaning of information for others	12.8	8.75	81.5
Cashiers	18	Performing administrative activities	68.1	3.18	81.5
First-line supervisors of retail sales workers	13	Guiding, directing, and motivating subordinates	19.7	15.02	77.4
Industrial engineers	20	Getting information	8.0	20.60	69.4
Customer service representatives	51	Performing for or working directly with the public	6.9	9.35	67.4
Lawyers	31	Providing consultation and advice to others	2.3	41.14	61.8
First-line supervisors of helpers, laborers, and material movers	24	Organizing, planning, and prioritizing work	8.5	12.73	54.2

1 Detailed work activity. There are 37 total DWA groups.

SOURCE: National labor and statistical sources; McKinsey Global Institute analysis

First, deep learning models are opaque, which can be a barrier to adoption in certain applications. As of today, it is difficult to decipher how deep neural networks reach insights and conclusions, making their use challenging in cases where transparency of decision making may be needed for regulatory purposes. Also, decision makers and customers may not buy into insights generated in a non-transparent way, especially when those insights are counterintuitive.

Second, there are ethical questions surrounding machine intelligence. One set of ethical concerns relates to real-world biases that might be embedded into training data. Another question involves deciding whose ethical guidelines will be encoded in the decision making of intelligence and who is responsible for the algorithm's conclusions. Leading artificial intelligence experts, through OpenAI, the Foundation for Responsible Robotics, and other efforts, have begun tackling these questions.

Third, the potential risks of labor disruption from the use of deep learning to automate activities are generating anxiety. There is historical precedent for major shifts among sectors and changes in the nature of jobs in previous waves of automation. In the United States, the share of farm employment fell from 40 percent in 1900 to 2 percent in 2000; similarly, the share of manufacturing employment fell from roughly 25 percent in 1950 to less than 10 percent in 2010. In both circumstances, while some jobs disappeared, new ones were created, although what those new jobs would be could not be ascertained at the time. But history does not necessarily provide assurance that sufficient numbers of new, quality jobs will be created at the right pace. At the same time, many countries have or will soon have labor forces that are declining in size, requiring an acceleration of productivity to maintain anticipated rates of economic growth. But automation technologies will not be widely adopted overnight; in fact, a forthcoming MGI research report will explore the potential pace of automation of different activities in different economies. Certainly dealing with job displacement, retraining, and unemployment will require a complex interplay of government, private sector, and educational and training institutions, and it will be a significant debate and an ongoing challenge across society.

•••

Data and analytics have even greater potential to create value today than they did when companies first began using them. Organizations that are able to harness these capabilities effectively will be able to create significant value and differentiate themselves, while others will find themselves increasingly at a disadvantage.

© Cultura Creative/Alamy

© Real444/Getty Images

1. THE DATA AND ANALYTICS REVOLUTION GAINS MOMENTUM

Back in 2011, the McKinsey Global Institute published a report highlighting the transformational potential of big data and analytics.[11] Five years later, we remain of the firm view that this potential has not been overhyped. In fact, we now believe that our 2011 analyses revealed only the tip of the iceberg. Today the range of applications and the opportunities associated with data and analytics have grown much larger.

The pace of change is accelerating. The volume of data continues to double every three years as information pours in from digital platforms, wireless sensors, virtual reality applications, and billions of mobile phones. Data storage capacity has increased, while its cost has plummeted. Data scientists now have unprecedented computing power at their disposal, and they are devising ever more sophisticated algorithms. The convergence of these trends is setting off industry disruptions—and posing new challenges for organizations.

Many companies have already harnessed these capabilities to improve their core operations or to launch entirely new business models. UPS feeds data into its ORION platform to determine the most efficient routes for its drivers dynamically. In the United States alone, the company estimates that the system will reduce the number of miles its vehicles travel each year by 100 million, saving more than $300 million annually.[12] Google is running a vast number of experiments to induce faster search queries, since a few milliseconds can translate into additional millions of dollars in revenue.[13] By merging information gleaned from social media with its own transaction data from customer relationship management and billing systems, T-Mobile US is reported to have cut customer defections in half in a single quarter.[14] Netflix has refined its recommendation engine and rolled it out to global customers—and a study released by the company estimates that it brings in $1 billion in annual revenue.[15]

But most companies remain in the starting gate. Some have invested in data and analytics technology but have yet to realize the payoff, while others are still wrestling with how to take the initial steps.

Digital native companies, on the other hand, were built for data and analytics–based disruption from their inception. It is easier to design new IT systems and business processes from scratch than to modify or overhaul legacy systems, and the top analytics talent tends to flock to organizations that speak their language. All of these advantages underscore the fact that incumbents need to stay vigilant about competitive threats and take a big-picture view of which parts of their business model are most vulnerable. In sectors as varied as transportation, hospitality, and retail, data and analytics assets have helped digital natives circumvent traditional barriers to entry (such as physical capital investments) and erect new ones (digital platforms with powerful network effects).

---

11 *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, June 2011.

12 "UPS ORION to be deployed to 70 percent of US routes in 2015; delivers significant sustainability benefits," company press release, March 2, 2015; Steven Rosenbush and Laura Stevens, "At UPS, the algorithm is the driver," *The Wall Street Journal*, February 16, 2015.

13 Jacques Bughin, "Big data, big bang?" *Journal of Big Data*, volume 3, number 2, 2016.

14 Paul Taylor, "Crunch time for big data," *The Financial Times*, June 19, 2012.

15 Carlos A. Gomez-Uribe and Neil Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems*, volume 6, number 4, January 2016.

We have reached a unique moment in time. Technology has begun to set off waves of accelerating change that will become the norm. Innovations in machine learning and deep learning have expanded the possibilities associated with big data well beyond what we foresaw in 2011.

ADVANCES IN DATA COLLECTION, MACHINE LEARNING, AND COMPUTATIONAL POWER HAVE FUELED PROGRESS

The sheer volume of available data has grown exponentially over the past five years, and new tools have been developed for turning this flood of raw data into insights. Machine learning, a term that encompasses a range of algorithmic approaches from statistical methods like regressions to neural networks, has rapidly advanced to the forefront of analytics. Underpinning all of this progress has been steady improvement in computational power from better processors and graphics processing units, combined with increased investment in massive computing clusters, often accessed as cloud services. The most sophisticated innovations are emerging from data scientists working at the intersection of academia and a few leading companies.

Each of these trends—more data, more tools for analyzing the data, and the firepower needed to do so—constitutes a breakthrough in its own right. Together these developments reinforce each other, and the broader field of big data and advanced analytics is making rapid advances as they converge. Vast increases in data require greater computational power and infrastructure to analyze and access them. Both data and computational power enable next-generation machine learning methods such as deep learning, which employs a deep graph of multiple processing layers that needs to be fed large quantities of data to produce meaningful insights. The confluence of data, storage, algorithms, and computational power today has set the stage for a wave of creative destruction.

The volume of available data has grown exponentially

For much of human history, the volume of available data grew slowly and in tandem with population growth. Only in the past few decades has this growth taken off as our methods for generating, collecting, and storing data have fundamentally changed. Proliferating data sources, from internet searches to social media activity to online transactions to sensors, are creating torrents of information. In fact, because so much of the world is instrumented, it is actually difficult to *avoid* generating data.

2x
growth in the volume of data expected every 3 years

As our previous report noted, just three exabytes of data existed in 1986—but by 2011, that figure was up to more than 300 exabytes. The trend has not only continued but has accelerated since then. One analysis estimates that the United States alone has more than two zettabytes (2,000 exabytes) of data, and that volume is projected to double every three years.[16]

Billions of people worldwide have gradually shifted more of their lives online, using their smartphones as ever-present personal command centers. A recent US survey found that 42 percent of Americans use the internet several times a day, and 21 percent report being online "almost constantly."[17] All of this activity leaves a rich data trail at every turn—in fact, this is one of the developments that gave rise to the term "big data." Internet companies and websites capture online behavior, including every web search, where people click, how long they stay on each site, and every transaction they make.

Internet users have gone from being passive consumers to active creators of content through social media and other forms—and digital native companies have capitalized by capturing these online data. More traditional businesses that rely on physical assets have

[16] John Gantz and David Reinsel, *The digital universe in 2020*, IDC, February 2013.

[17] Andrew Perrin, *One-fifth of Americans report going online "almost constantly,"* Pew Research Center, December 2015 (survey conducted July 2015).

been slower to realize gains from data and analytics, but as they continue to digitize their customer-facing and internal operations, they, too, now have the building blocks for more sophisticated use of data and analytics.

Data has not only increased in volume; it has also gained tremendous richness and diversity. We have entered a new era in which the physical world has become increasingly connected to the digital world. Data is generated by everything from cameras and traffic sensors to heart rate monitors, enabling richer insights into human behavior. Retailers, for instance, are trying to build complete customer profiles across all their touch points, while health-care providers can now monitor patients remotely. In addition, companies in traditional industries such as natural resources and manufacturing are using sensors to monitor their physical assets and can use those data to predict and prevent bottlenecks or to maximize utilization.

Much of this newly available data is in the form of clicks, images, text, or signals of various sorts, which is very different than the structured data that can be cleanly placed in rows and columns. Additional storage technologies, including non-relational databases, have allowed businesses to collect this rich content. A great deal of the potential it holds is still unrealized.[18] But new tools and applications of data and analytics could eventually transform even traditional industries that once seemed far removed from the digital world.

Algorithms have continued to advance

The increasing availability of data has fueled advances in analytical techniques and technologies, with machine learning at the forefront.

A standard software program is hard-coded with strict rules for the tasks it needs to execute. But it cannot adapt to new variables or requirements unless a programmer updates it with specific new rules. While this works well in some contexts, it is easy to see why this approach is not scalable to handle all the complexities of the real world. Machine learning, meanwhile, uses an inductive approach to form a representation of the world based on the data it sees. It is able to tweak and improve its representation as new data arrive. In that sense, the algorithm "learns" from new data inputs and gets better over time. The key requirement for machine learning is vast quantities of data, which are necessary to train algorithms. Vastly larger quantities of rich data have enabled remarkable improvements in machine learning algorithms, including deep learning.

Among the most important advances in machine learning techniques over the past few years are the following:

- **Deep learning.** This branch of machine learning uses deep neural networks with many hidden layers. Two of the most common types of deep neural networks are convolutional and recursive. Convolutional neural networks are often used for recognizing images by processing a hierarchy of features—for instance, making the connection between a nose, a face, and eventually a full cat. This image recognition capability has important applications in the development of autonomous vehicles, which need to recognize their surroundings instantly. In contrast, recursive neural networks are used when the overall sequence and context are important, as in speech recognition or natural language processing.

  Deep learning systems are the clearest example of the confluence of abundant data, processing power, and increasingly sophisticated algorithms. Neural networks were developed decades ago, but they lacked the massive quantities of data and processing power needed to reach their full capabilities. Now that those barriers have been

[18] *The internet of things: Mapping the value beyond the hype,* McKinsey Global Institute, June 2015.

overcome, data scientists are making rapid advances in deep learning techniques. These will be discussed more fully in Chapter 5.

- **Reinforcement learning.** Another form of machine learning, reinforcement learning, takes actions toward a specified goal, but without direction on which actions to take in order to get there. The algorithms explore a broad range of possible actions while gradually learning which ones are most effective, thereby incorporating an element of creativity. Reinforcement learning has been applied to applications ranging from learning how to play games like chess and Go to improving traffic management at stoplights.[19]

- **Ensemble learning.** This set of techniques uses multiple machine learning methods to obtain better predictions than any one method could achieve on its own. Ensemble methods are particularly useful when there is a wide range of possible hypotheses, as they help to zero in on the most appropriate path. CareSkore, for example, is employing ensemble learning using Google's TensorFlow platform to analyze a range of sociodemographic and behavioral data with the goal of improving preventive health care.

These new techniques are made possible by new tools. Deep learning libraries and platforms such as TensorFlow, Caffe, and Theano allow practitioners to integrate deep learning algorithms into their analysis quickly and easily. Spark offers a big data platform that provides advanced real-time and predictive analytics applications on widely used Apache Hadoop distributed storage systems. New machine learning application program interfaces (APIs), such as Microsoft's ML API, enable users to implement machine learning in new areas.

Greater computational power and storage capacity enable greater use of data and analytics

40x
increase in processing power between fastest supercomputer of 2010 and the fastest today

Over the past five years, computational power has continued to grow, reflecting the continuing explanatory power of Moore's Law. In 1993, computers could handle 4x10^9 instructions per second, and by the time our 2011 report was published, that figure had grown by more than three orders of magnitude, to 6x10^12 instructions per second.[20] Meanwhile, the amount of computing power each dollar can buy has increased by a factor of ten roughly every four years in the past quarter century, making cheap computing more available than ever before.[21] Continued investment is pushing the boundaries of these capabilities. In 2016, China unveiled the world's fastest supercomputer, which is more than 40 times as powerful as the fastest computer of 2010.[22]

Even as Moore's Law is nearing its physical limitations, other innovations are fueling continued progress. Computational power has gotten a boost from an unlikely source: video game consoles. The need for computational power to power ever more sophisticated video games has spurred the development of graphics processing units. GPUs have enabled image processing in neural networks that is ten to 30 times as fast as what conventional CPUs can do. These processors have been put to use in many non-gaming contexts, ranging from Cortexica's image recognition programs to Salesforce.com's Twitter analysis.[23]

---

19 Lior Kuyer et al., "Multiagent reinforcement learning for urban traffic control using coordination graphs," *Machine learning and knowledge discovery in databases*, volume 5211 of Lecture Notes in Computer Science series, Springer Berlin Heidelberg, 2008.

20 Martin Hilbert and Priscila López, "The world's technology capacity to store, communicate, and compute information," *Science*, volume 332, issue 6025, April 2011.

21 Luke Muehlhauser and Lila Rieber, "Exponential and non-exponential trends in information technology," Machine Intelligence Research Institute blog, May 12, 2014.

22 Top500 list of world's supercomputers, June 2016, available at https://www.top500.org/list/2016/06/.

23 "Shazam, Salesforce.com, Cortexica use GPUs for affordable, scalable growth in consumer and commercial applications," Nvidia corporate press release, March 2013.

The existence of so much computational power in many hands has made new forms of crowdsourcing possible though distributed computing, which takes advantage of a large network of idle computers. Extremely time-consuming simulations of protein molecules for medical research have been performed across a large number of volunteer computers through Folding@home, a distributed computing project based out of Stanford University's Pande Lab, for example.[24]

The growth of cloud-based platforms has given virtually any company the tools and storage capacity to conduct advanced analytics. Cloud-based storage and analytics services enable even small firms to store their data and process it on distributed servers. Companies can purchase as much space as they need, greatly simplifying their data architecture and IT requirements and lowering capital investment. As computation capacity and data storage alike have largely been outsourced, many tools have become accessible, and data can now be more easily combined across sources. NoSQL databases offer alternatives to relational databases, allowing for the collection and storage of various types of unstructured data, such as images, text, audio, and other rich media.

Data storage costs have fallen dramatically over the years. But now the increasing firehose of data streams simply exceeds the amount of storage that exists in the world—and projections indicate that the production of new storage capacity will continue to fall well short of the demand created by explosive growth in data generation.[25] Previous MGI research found that more than 40 percent of all data generated by sensors on the typical oil rig were never even stored. This flood of data has to be captured and retained to be easily accessed for analysis before it can yield value.

ANALYTICS CAPABILITIES ARE ALREADY LEADING TO NEW BUSINESS MODELS AND RESHAPING INDUSTRY COMPETITION

Companies at the forefront of the technology trends described above are using their capabilities to establish formidable advantages. We now see huge disparities in performance between a small group of leading companies and the average company. Some markets now have winner-take-most dynamics. Analytics capabilities have become a differentiating factor in industry competition, as leading players use data and analytics to grow revenue, to enter or even create new markets, to change the nature of their relationship with customers, and to increase organizational efficiencies. Organizations that are lagging behind will need to adapt quickly before the gap grows wider.

Data has become the new corporate asset class—and the best way for companies to generate and access it is to digitize everything they do. Digitizing customer interactions provides a wealth of information for marketing, sales, and product development, while internal digitization generates data that can be used to optimize operations and improve productivity.

Looking at a wide variety of indicators that measure digitization, we see a striking gap—particularly in digital usage and labor—between leading firms and the laggards. Digitization tends to be correlated with data and analytics capabilities since it is a crucial enabler of data generation and collection; it also requires related skills and mindsets as well as process and infrastructure changes. Of the domains we reviewed in our 2011 research, information and communications technology (the sector that has made the greatest advances in location-based data), retail leaders, and advanced manufacturing rank among the leaders in MGI's Industry Digitization Index. Meanwhile, government, basic manufacturing, the bulk of the

[24] Vijay Pande, "FAH's achievements in 2015, with a glimpse into 2016," Folding@home blog, December 6, 2015.

[25] See, for example, Drew Robb, "Are we running out of data storage space?" World Economic Forum blog, November 10, 2015, as well as *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, June 2011.

retail sector, and health care are lagging behind in measures of digitization—and in analytics usage.[26] If companies and organizations in those sectors continue to lag in the pace of adoption, they could open the door for new digital native challengers.

Analytics leaders such as Apple, Alphabet/Google, Amazon, Facebook, Microsoft, GE, Baidu, Alibaba Group, and Tencent have established themselves as some of the most valuable companies in the world.[27] These companies have differentiated themselves through unique data sources, a wealth of analytics talent, and investment in data infrastructure.[28] The same trend can be seen among the next wave of disruptors: the privately held global "unicorns." These tend to be companies with business models predicated on data and analytics, such as Uber, Lyft, Didi Chuxing, Palantir, Flipkart, Airbnb, DJI, Snapchat, Pinterest, BlaBlaCar, Ola, Snapdeal, and Spotify. This list shows that the next generation of digital leaders is becoming more global. Data and analytics capabilities have moved beyond the Western world toward Asia, especially India and China.

Data permeates everything that the leading organizations do. Digitizing customer interactions provides a wealth of information that can feed into strategy, marketing, sales, and product development. Increasingly granular data allows companies to micro-target new customers to acquire and to develop products and services that are more personalized. Internal digitization generates data that can be used to make operations more efficient, including better sourcing, supply-chain and logistics management, and predictive maintenance on equipment.

The value of data and analytics has upended the traditional relationship between consumers and producers. In the past, companies sold products to their customers in return for money and negligible data. Today, transactions—and indeed every interaction with a consumer—produce valuable information. Sometimes the data itself is so valuable that companies such as Facebook, LinkedIn, Pinterest, Twitter, and many others are willing to offer free services in order to obtain it. In some cases, the "customer" is actually a user who barters his or her data in exchange for free use of the product or service. In others, the actual customers may be marketers that pay for targeted advertising based on user-generated data. To maintain an edge in consumer data, user acquisition and user interaction are both critical. Venture capitalists have long understood the importance of building a customer base. Many internet startups from Quora to Jet have focused as much attention on capturing users who can provide valuable data as on capturing paying customers.[29]

These technologies are allowing new players to challenge incumbents with surprising speed since they circumvent the need to build traditional fixed assets. Amazon, Netflix, Uber, Airbnb, and a host of new "fintech" financial firms have moved into industries where incumbents were heavily invested in certain types of physical assets. These disrupters used their digital, data, and analytics assets to create value without owning physical shops, cable connections to viewers' homes, car fleets, hotels, or bank branches, respectively. But even as they bypassed traditional barriers to entry, they have erected new ones. The network effects of digital marketplaces, social networks, and other digital platforms can create a winner-take-most phenomenon. The leading platforms capture a disproportionate share of the data created in a given space, making it difficult for new entrants to challenge them.

---

[26] *Digital America: A tale of the haves and have-mores*, McKinsey Global Institute, December 2015; and *Digital Europe: Pushing the frontier, capturing the benefits*, McKinsey Global Institute, June 2016.

[27] Michael Chui and James Manyika, "Competition at the digital edge: 'Hyperscale' businesses," *McKinsey Quarterly*, March 2015.

[28] Jacques Bughin, "Big data: Getting a better read on performance," *McKinsey Quarterly*, February 2016.

[29] Maya Kosoff, "These 15 startups didn't exist five years ago—now they're worth billions," *Business Insider*, December 10, 2015.

The leading firms have a remarkable depth of analytical talent deployed on a variety of problems—and they are actively looking for ways to make radical moves into other industries. These companies can take advantage of their scale and data insights to add new business lines, and those expansions are increasingly blurring traditional sector boundaries.[30] Alphabet, which used its algorithmic advantage to push Google ahead of older search engine competitors, is now using its formidable talent base to expand into new business lines such as the development of autonomous vehicles. Apple used its unique data, infrastructure edge, and product platform to push into the world of finance with Apple Pay. Similarly, Chinese e-commerce giants Alibaba, Tencent, and JD.com have leveraged their data volumes to offer microloans to the merchants that operate on their platforms. By using real-time data on the merchants' transactions to build its own credit scoring system, Alibaba's finance arm was able to achieve better non-performing loan ratios than traditional banks.[31] Furthermore, banks and telecom companies are sharing data to drive new products and to improve core operations such as credit underwriting, customer segmentation, and risk and fraud management.

To keep up with the pace of change, incumbents need to look through two lenses at the same time. In an environment of constant churn, organizations should have one eye on high-risk, high-reward moves of their own. They have to keep thinking ahead, whether that means entering new markets or changing their business models. At the same time, they have to maintain a focus on using data and analytics to improve their core business. This may involve identifying specific use cases and applications on the revenue and the cost side, using analytics insights to streamline internal processes, and building mechanisms for constant learning and feedback to improve. Pursuing this two-part strategy should position any organization to take advantage of opportunities and thwart potential disruptors. See Chapter 2 for further discussion of the steps involved in transforming a traditional company into a more data-driven enterprise.

•••

With data and analytics technology rapidly advancing, the next question is how companies will integrate these capabilities into their operations and strategies—and how they will position themselves in a world where analytics capabilities are rapidly reshaping industry competition. But adapting to an era of more data-driven decision making is not always a simple proposition for people or organizations. Even some companies that have invested in data and analytics capabilities are still struggling with how to capture value from the data they are gathering. The next chapter revisits our 2011 research. It estimates how much progress has been made in the five domains where we previously highlighted tremendous potential—and points to even wider opportunities that exist today to create value through analytics.

[30] *Playing to win: The new global competition for corporate profits*, McKinsey Global Institute, September 2015.

[31] *China's digital transformation: The internet's impact on productivity and growth*, McKinsey Global Institute, July 2014.

© PhotoAlto/Odilon Dimier/Getty Images

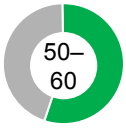
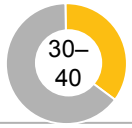
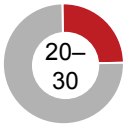
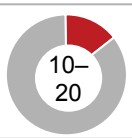
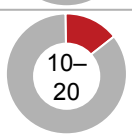
2. OPPORTUNITIES STILL UNCAPTURED

This chapter revisits the five domains we highlighted in our 2011 report to evaluate their progress toward capturing the potential value that data and analytics can deliver. In each of these areas, our previous research outlined dozens of avenues for boosting productivity, expanding into new markets, and improving decision making. Below we examine how much of that value is actually being captured, combining quantitative data with input from industry experts. The numerical estimates of progress in this chapter provide an indication of which areas have the greatest momentum and where barriers have proven to be more formidable, although we acknowledge they are directional rather than precise.

We see the greatest progress in location-based services and in retail, where competition from digital native firms has pushed other players toward adoption (Exhibit 1). The thin margins facing retailers (especially in the grocery sector) and pressure from competitors such as Amazon and leading big-box retailers such as Walmart create a strong incentive to evolve. In contrast, manufacturing, the public sector, and health care have captured less than a third of the value opportunities that data and analytics presented five years ago. Overall, many of the opportunities described in our 2011 report are still on the table. In the meantime, the potential for value creation has grown even bigger.

Exhibit 1

There has been uneven progress in capturing value from data and analytics

	Potential impact: 2011 research	Value captured %	Major barriers
Location-based data	▪ $100 billion+ revenues for service providers ▪ Up to $700 billion value to end users		▪ Penetration of GPS-enabled smartphones globally
US retail[1]	▪ 60%+ increase in net margin ▪ 0.5–1.0% annual productivity growth		▪ Lack of analytical talent ▪ Siloed data within companies
Manufacturing[2]	▪ Up to 50% lower product development cost ▪ Up to 25% lower operating cost ▪ Up to 30% gross margin increase		▪ Siloed data in legacy IT systems ▪ Leadership skeptical of impact
EU public sector[3]	▪ ~€250 billion value per year ▪ ~0.5% annual productivity growth		▪ Lack of analytical talent ▪ Siloed data within different agencies
US health care	▪ $300 billion value per year ▪ ~0.7% annual productivity growth		▪ Need to demonstrate clinical utility to gain acceptance ▪ Interoperability and data sharing

1 Similar observations hold true for the EU retail sector.
2 Manufacturing levers divided by functional application.
3 Similar observations hold true for other high-income country governments.

SOURCE: Expert interviews; McKinsey Global Institute analysis

Most business leaders recognize the size of the opportunities and feel the pressure to evolve. Recent research has found that investing in data and analytics capabilities has high returns, on average: firms can use these capabilities to achieve productivity gains of 6 to 8 percent, which translates into returns roughly doubling their investment within a decade. This is a higher rate of return than other recent technologies have yielded, surpassing even the computer investment cycle in the 1980s.[32]

However, these high returns are largely driven by only a few successful organizations. Early adopters are posting faster growth in operating profits, which in turn enables them to continue investing in data assets and analytics capabilities, solidifying their advantages. Facebook, in particular, has created a platform capable of gathering remarkably detailed data on billions of individual users. But not all of the leaders are digital natives. Walmart, GE, Ferrari F1, and Union Pacific are examples of companies in traditional industries whose investments in data and analytics have paid significant dividends on both the revenue and cost sides.

Many other companies are lagging behind in multiple dimensions of data and analytics transformation—and the barriers are mostly organizational issues. The first challenge is incorporating data and analytics into a core strategic vision. The next step is developing the right business processes and building capabilities (including both data infrastructure and talent); it is not enough to simply layer powerful technology systems on top of existing business operations. All these aspects of transformation need to come together to realize the full potential of data and analytics—and the challenges incumbents face in pulling this off are precisely why much of the value we highlighted in 2011 is still unclaimed.

LOCATION-BASED SERVICES HAVE MADE THE MOST SIGNIFICANT PROGRESS SINCE 2011

Location-based services, which exist across multiple industries, use GPS and other data to pinpoint where a person (or a vehicle or device) is situated in real time. This domain has made the greatest strides since our 2011 report, thanks in large part to the widespread adoption of GPS-enabled smartphones. The role of digital giants such as Google and Apple in driving these applications forward for billions of smartphone users is hard to overstate.

We estimate that some 50 to 60 percent of the potential value anticipated in our 2011 research from location-based services has already been captured. We looked separately at the revenue generated by service providers and the value created for consumers. Our 2011 report estimated that service providers had roughly $96 billion to $106 billion in revenue at stake from three major sources: GPS navigation devices and services, mobile phone location-based service applications, and geo-targeted mobile advertising services. Today's market has already reached 60 percent of that revenue estimate, with particularly strong growth in the use of GPS navigation. Industries and consumers alike have embraced real-time navigation, which is now embedded in a host of services that monetize this technology in new ways. Uber and Lyft use location data for their car dispatching algorithms, and online real estate platforms such as Redfin embed street views and neighborhood navigation in their mobile apps to aid home buyers.

Our 2011 analysis estimated that end consumers would capture the equivalent of more than $600 billion in value, which is the lion's share of the benefits these services create. The world is at the tipping point at which smartphones account for most mobile phone subscriptions (although most people in the world still do not own mobile phones).[33] The share is much higher in advanced economies and is rising rapidly worldwide. This trend puts mapping

[32] Jacques Bughin, "Ten lessons learned from big data analytics," *Journal of Applied Marketing Analytics*, forthcoming.

[33] 451 Research data. See also *Ericsson mobility report: On the pulse of the networked society*, Ericsson, June 2016.

technology in the pockets of billions of consumers. Much of the value comes, as expected in 2011, in the form of time and fuel savings as they use GPS navigation while driving and adopt many other mobile location-based services that deliver new types of conveniences.

Yet the opportunities have grown beyond what we envisioned in our 2011 report. Today there are new and growing opportunities for businesses in any industry to use geospatial data to track assets, teams, and customers in dispersed locations in order to generate new insights and improve efficiency. These opportunities are significant and, while still in the very early stages, could turn out to be even larger than the ones discussed above.

LEADERS IN THE RETAIL SECTOR HAVE ADOPTED ANALYTICS, BUT MARGINS REMAIN THIN AS MUCH OF THE VALUE GOES TO CONSUMER SURPLUS

Analytics has tremendous relevance for retailers, since they can mine a trove of transaction-based and behavioral data from their customers. In 2011, we estimated that some retailers could increase net margins by more than 60 percent—and that the US sector as a whole could boost annual productivity growth by 0.5 to 1.0 percent. Five years later, US retailers have captured 30 to 40 percent of this potential.

30–40%
potential value captured in the retail sector

While our 2011 analysis focused on the US sector alone, these opportunities clearly exist in other high-income countries as well, especially in Europe. Moreover, the incentives for adoption are there; major retailers worldwide have been early adopters of analytics as they seek to respond to the competitive pressures created by e-commerce. But despite the improvements made possible by analytics, overall margins have remained thin (the earnings before interest, taxes, and amortization, or EBITA, margins held steady around 7 to 8 percent from 2011 to 2016).[34] This is because a great deal of the value has gone to consumers, who have been the major beneficiaries of intense competition in the retail world.

Capabilities are also uneven across the industry. Big-box retailers such as Target, Best Buy, and Costco have invested in creating an end-to-end view of their entire value chain, from suppliers to warehouses to stores to customers. Real-time information from its stores allowed Globus, a department store chain in Switzerland, to update its product mix and respond quickly to customer demand.[35] In addition, certain subsectors have made faster progress. The grocery sector has led the way, while smaller retailers specializing in clothing, furnishings, and accessories have lagged behind. Organizational hurdles, including the difficulty of finding data scientists and breaking down information silos across large companies, have kept many companies from realizing the full potential payoff.

Our 2011 report focused on integrating analytics into five key functions: marketing, merchandising, operations, supply-chain management, and new business models. We have seen fairly even progress across all of these, with most large retailers adopting at least basic analytics to optimize their operations and supply chains.

In marketing and sales, the biggest emphasis has been on improved cross-selling, including "next product to buy" recommendations. While Amazon pioneered this use of technology, many other retailers (including Nordstrom, Macy's, and Gilt) now make recommendations based on user data. In addition, retailers have tested everything from location-based ads to social media analysis; Target, for instance, is piloting the use of beacons that transmit targeted in-store ads depending on a shopper's precise position.[36] Within merchandising, retailers have made strides in optimizing their pricing (especially online) and assortment, but they have not brought as much technology to bear on placement. Amazon is mining

[34] Based on a sample of a dozen large global retail companies from their annual reports.

[35] *Real-time enterprise stories*, case studies from Bloomberg Businessweek Research Services and Forbes Insights, SAP, October 2014.

[36] Sarah Perez, "Target launches beacon test in 50 stores, will expand nationwide later this year," TechCrunch, August 5, 2015.

other sellers on its marketplace for possible additions to its own assortment. Among the operations and supply-chain levers, there has been widespread adoption for uses such as making performance more transparent, optimizing staffing levels, improving sourcing, and streamlining shipping and transport.

PROGRESS IN MANUFACTURING HAS LARGELY BEEN LIMITED TO A SMALL GROUP OF INDUSTRY LEADERS

20-30%
potential value captured in the manufacturing sector

Manufacturing industries have captured only about 20 to 30 percent of the potential value we estimated in our 2011 research—and most of that has gone to a handful of industry-leading companies. Those that made decisive investments in analytics capabilities have often generated impact in line with our estimates. The sector's main barrier seems to be the perception of many companies that the complexity and cost of analytics could outweigh the potential gains, particularly if the companies have difficulty identifying the right technology and talent. There is no single integrated system that is a clear choice for every company. Many will not solve the full problem of data being cordoned off in silos across an organization, and installing replacement systems is a difficult undertaking.

Our 2011 report highlighted opportunities for the global manufacturing sector to realize value from data and analytics within R&D, supply-chain management, production, and after-sales support. Within R&D, design-to-value (the use of customer and supplier data to refine existing designs and feed into new product development) has had the greatest uptick in adoption, particularly among carmakers. While adoption of advanced demand forecasting and supply planning has been limited, there are some individual success stories. One stamping parts producer was able to save approximately 15 percent on product costs by using these types of insights to optimize its production footprint.

Within the actual production process, the greatest advances have been in developing digital models of the entire production process. Industry leaders such as Siemens, GE, and Schneider Electric have used these "digital factories" to optimize operations and shop floor layout, though this technique often focuses on designing new facilities. Furthermore, throughout ongoing production processes, many early adopters are using sensor data to reduce operating costs by some 5 to 15 percent. Data-driven feedback in after-sales has been most heavily applied within servicing offers, especially in aerospace or large installations in business-to-business transactions. After-sales servicing offers once relied on simple monitoring, but now they are beginning to be based on real-time surveillance and predictive maintenance.

MUCH OF THE POTENTIAL FOR COST SAVINGS, EFFICIENCY, AND TRANSPARENCY IN THE PUBLIC SECTOR REMAINS UNREALIZED

10-20%
potential value captured by the EU public sector

Our 2011 report analyzed the public sector in the European Union (EU), where we outlined some €250 billion worth of annual savings that could be achieved by making government services more efficient, reducing fraud and errors in transfer payments, and improving tax collection. But little of this has materialized, as EU agencies have captured only about 10 to 20 percent of this potential.

In terms of operations, some government entities have moved more interactions online, and many (particularly tax agencies) have adopted more pre-filled forms. There is also a movement to improve data sharing across agencies, exemplified in the EU initiative called "Tell It Once." On a country-specific level, the Netherlands has moved most tax and social welfare functions online, while France saved the equivalent of 0.4 percent of GDP by reducing requests from agencies to citizens for the same type of information from 12 to one.

Adoption of algorithms to detect fraud and errors in transfer payments has been limited. Analytics have been used to improve the rate of tax collection, mainly by targeting tax audits more effectively and running algorithms on submitted tax forms. France has automated

cross-checks between agencies to improve the accuracy of their reviews of tax forms, while the United Kingdom's payment phase segmentation leading to targeted tax audits recovered some £2 billion in its first year.

While our 2011 report focused on the EU public sector, these observations regarding government adoption are applicable across all high-income economies. Adoption and capabilities generally vary greatly from country to country, and even among agencies (with tax agencies typically being the most advanced within a given country). The main barriers holding back progress have been organizational issues, the deeply entrenched nature and complexity of existing agency systems, and the difficulty of attracting scarce analytics talent with public-sector salaries.

US HEALTH-CARE PLAYERS HAVE ONLY BEGUN TO DEVELOP ANALYTICS CAPABILITIES AND TRANSFORM THE DELIVERY OF PATIENT CARE

10–20%
potential value captured by the US health-care sector

Our 2011 report outlined $300 billion worth of value that big data analytics could unlock in the US health-care sector. To date, only 10 to 20 percent of this value has been realized. Making a major shift in how data is used is no easy task in a sector that is not only highly regulated but often lacks strong incentives to promote increased usage of analytics. A range of barriers—including a lack of process and organizational change, a shortage of technical talent, data-sharing challenges, and regulations—have combined to limit the impact of data and analytics throughout the sector and constrain many of the changes we envisioned.

The opportunities we highlighted were split among five categories: clinical operations, accounting and pricing, R&D, new business models, and public health. Within clinical operations, the major success has been the rapid expansion of electronic medical records (EMRs), which accounted for 15.6 percent of all records in 2010 but more than 75 percent by 2014, aided by the incentives for providers in the Affordable Care Act.[37] This has enabled basic analytics but little has been done to unlock and fully utilize the vast stores of data actually contained within EMRs. A few providers have pushed this further, including Sutter Health, whose new EMR system processes reports 40 times faster and achieves an 87 percent increase in predicting readmissions compared with its previous system, by centralizing the data and analytics and pushing toward prospective analyses.[38]

Payers have also been slow to capitalize on big data for accounting and pricing, but a few encouraging trends have emerged. Transparency in health-care pricing has improved thanks to steps taken by the Centers for Medicare and Medicaid Services at the national level, while more than 30 states have established all-payer claims databases to serve as large-scale repositories of pricing information. A few insurers have made gains. Optum within UnitedHealth saves employers money by combing claim records for over-prescriptions.

Greater progress has been made in the pharmaceutical industry, where many companies have adopted analytics to assist their R&D, although they are still in the early stages of putting the full capabilities to work. Most pharma companies now use predictive modeling to optimize dosing as they move from animal testing to phase I clinical trials, but analytics have not yet been used as widely in later trials to determine questions such as the proper efficacy window and patient exclusion criteria. Data are being used in R&D to identify the right target population for drug development, which can reduce the time and cost of clinical trials by 10 to 15 percent. Contract research organizations, which are used more widely today than even five years ago, generally use statistical tools to improve the administration of clinical

[37] Dustin Charles, Meghan Gabriel, and Talisha Searcy, "Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2014," Office of the National Coordinator for Health Information Technology, data brief number 23, April 2015.

[38] "Healthcare providers unlock value of on-demand patient data with SAP HANA," SAP press release, May 4, 2016.

trials, and they are now using their scale to draw broader conclusions from their data. A few leading players have been aggressive in using clinical trial data to pursue label expansion opportunities (that is, finding additional uses for the same drug). Meanwhile, the Food and Drug Administration has introduced the Sentinel program to scan the records of 178 million patients for signs of adverse drug effects.

Some of the new models we highlighted in 2011 are in fact taking root. There is now a large and growing industry that aggregates and synthesizes clinical records. Explorys, for example, a data aggregator with some 40 million EMRs, was recently acquired by IBM to support the development of Watson. However, online platforms and communities (such as PatientsLikeMe) had strong initial success as key data sources, but other sources have also appeared. The use of analytics in public health surveillance has assumed new importance, given recent outbreaks of the Ebola and Zika viruses.

The health-care sector may have a long way to go toward integrating data and analytics. But in the meantime, the possibilities have grown much bigger than what we envisioned just five short years ago. Cutting-edge technology will take time to diffuse throughout the health-care system, but the use of machine learning to assist in diagnosis and clinical decision making has the potential to reshape patient care. Advances in deep learning in the near future, especially in natural language and vision, could help to automate many activities in the medical field, leading to significant labor cost savings. With labor constituting 60 to 70 percent of hospital total costs, this presents significant opportunities in the future.

1/10TH
the cost of sequencing a genome today as a share of the cost in 2011

The biggest frontier for data analytics in health care is the potential to launch a new era of truly personalized medicine (see Chapter 4 for a deeper discussion of what this could entail). New technologies have continued to push down the costs of genome sequencing from $10,000 in 2011 to approximately $1,000 today.[39] Combining this with the advent of proteomics (the study of proteins) has created a huge amount of new biological data. To date, the focus has been largely within oncology, as genomics has enabled characterization of the microsegments of each type of cancer.

A NUMBER OF BARRIERS STILL NEED TO BE OVERCOME

What explains the relatively slow pace of adoption and value capture in these domains and many others? Below we look at some of the internal and external barriers organizations face as they try to shift to a more data-driven way of doing business.

Adopting analytics is a multistep process within an organization

The relatively slow pace of progress in many domains points to some hurdles that most organizations encounter as they try to integrate analytics into their decision.

Many organizations have responded to competitive pressure by making large technology investments—without adopting the necessary organizational changes to make the most of them.

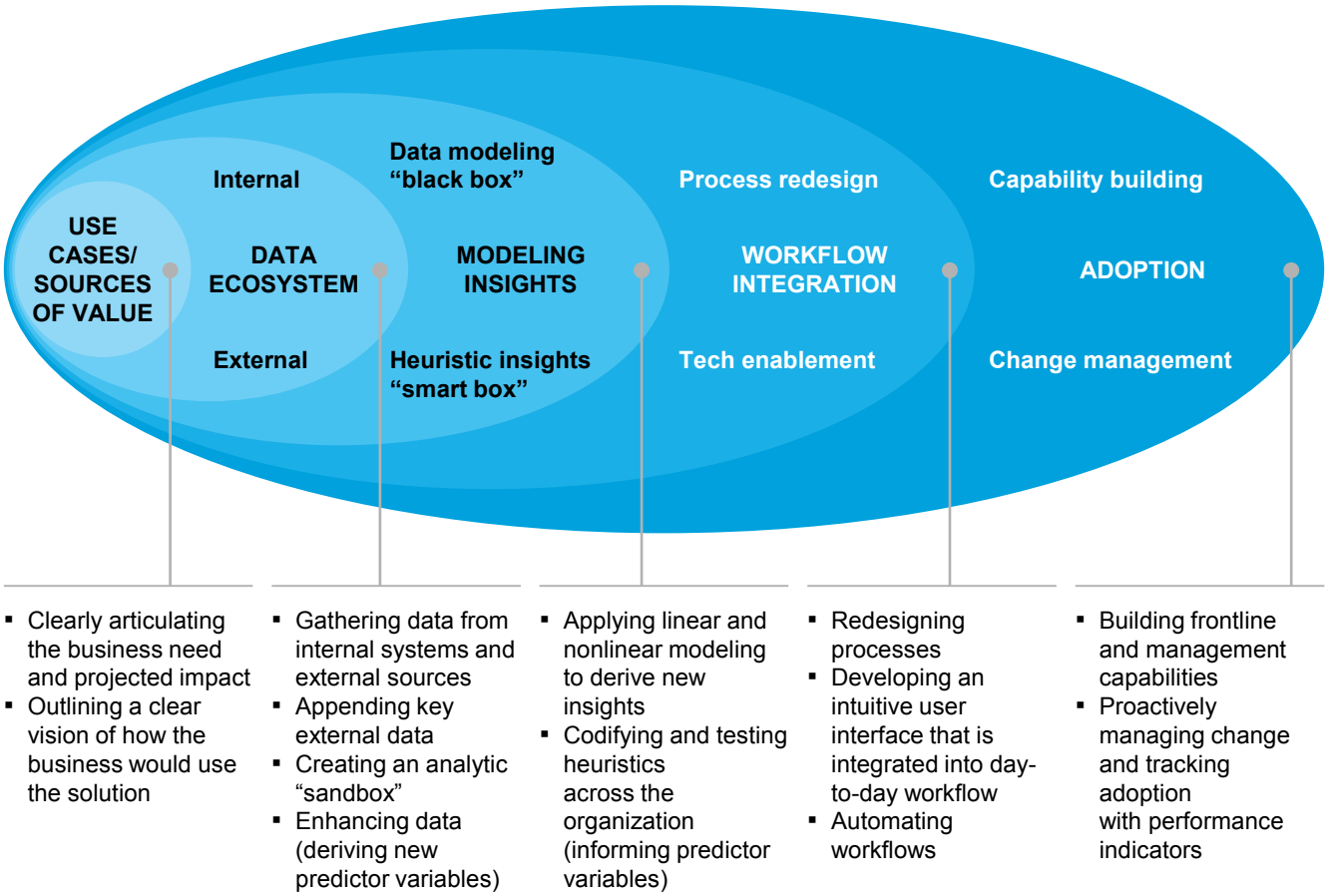
An effective transformation strategy can be broken down into several elements (Exhibit 2). The first is stepping back to ask some fundamental questions that can shape the strategic vision: What will data and analytics be used for? How will the insights drive value? How will the value be measured? The second component is building out the underlying data architecture as well as data collection or generation capabilities. Many incumbents struggle with switching from legacy data systems to a more nimble and flexible architecture to store and harness big data; they may also need to complete the process of fully digitizing transactions and processes in order to collect all the data that could be useful. The third element is acquiring the analytics capabilities needed to derive insights from data;

[39] Data from the NHGRI Genome Sequencing Program, National Human Genome Research Center, available at https://www.genome.gov/sequencingcostsdata/.

organizations may choose to add in-house capabilities or to outsource to specialists (see below for more on the talent shortage). The fourth element is another potential stumbling block: changing business processes to incorporate data insights into the actual workflow. This requires getting the right data insights into the hands of the right personnel within the organization. Finally, organizations need to build the capabilities of executives and mid-level managers to understand how to use data-driven insights—and to begin to rely on them as the basis for making decisions.

Exhibit 2

Successful data and analytics transformation requires focusing on five elements


USE CASES/ SOURCES OF VALUE
DATA ECOSYSTEM
Internal
External
MODELING INSIGHTS
Data modeling "black box"
Heuristic insights "smart box"
WORKFLOW INTEGRATION
Process redesign
Tech enablement
ADOPTION
Capability building
Change management
▪ Clearly articulating the business need and projected impact
▪ Outlining a clear vision of how the business would use the solution
▪ Gathering data from internal systems and external sources
▪ Appending key external data
▪ Creating an analytic "sandbox"
▪ Enhancing data (deriving new predictor variables)
▪ Applying linear and nonlinear modeling to derive new insights
▪ Codifying and testing heuristics across the organization (informing predictor variables)
▪ Redesigning processes
▪ Developing an intuitive user interface that is integrated into day-to-day workflow
▪ Automating workflows
▪ Building frontline and management capabilities
▪ Proactively managing change and tracking adoption with performance indicators

SOURCE: McKinsey Analytics; McKinsey Global Institute analysis

Failing to execute these steps well can limit the potential value. Digital native companies have a huge natural advantage in these areas. It is harder for traditional companies to overhaul or change existing systems, but hesitating to get started can leave them vulnerable to being disrupted. And while it may be a difficult transition, some long-established companies—including GE, Mercedes-Benz, Ferrari F1, and Union Pacific—have managed to pull it off. (See Box 1, "Identifying the most critical internal barriers for organizations.")

Box 1. Identifying the most critical internal barriers for organizations

McKinsey & Company recently conducted a survey of C-suite executives and senior managers regarding the use of data and analytics across a variety of industries.[1] The participants were asked which challenges were most pervasive in their company. Since respondents were asked to rank their top three challenges, the survey is not directly comparable between industries but only on the relative ranking of difficulties within an industry. However, it is useful for highlighting the barriers perceived by those in industries where progress has been slower (Exhibit 3). The barriers discussed in the survey can be broken into three categories: strategy, leadership, and talent; organizational structure and processes; and technology infrastructure.

Exhibit 3

Survey respondents report that strategic, leadership, and organizational hurdles often determine the degree to which they can use data and analytics effectively

Which of these have been among the TOP 3 most significant challenges to your organization's pursuit of its data and analytics objectives?

● High ● Moderate ● Low

	Barriers	Overall %	High tech and telecom	Retail	Manu-factur-ing	Public sector	Health care
Strategy, leadership, and talent	Constructing a strategy	30	●	●	●	●	●
	Ensuring senior management involvement	42	●	●	●	●	●
	Securing internal leadership for data and analytics projects	33	●	●	●	●	●
	Attracting and/or retaining appropriate talent (both functional and technical)	21	●	●	●	●	●
Organi-zational structure and processes	Tracking the business impact of data and analytics activities	23	●	●	●	●	●
	Designing an appropriate organizational structure to support data and analytics activities	45	●	●	●	●	●
	Creating flexibility in existing processes to take advantage of data-driven insights	13	●	●	●	●	●
IT infra-structure	Providing business functions with access to support	14	●	●	●	●	●
	Investing at scale	17	●	●	●	●	●
	Designing effective data architecture and technology infrastructure	36	●	●	●	●	●

SOURCE: McKinsey Global Institute analysis

[1] "The need to lead in data and analytics," McKinsey & Company survey, McKinsey.com, April 2016, available at http://www.mckinsey.com/business-functions. The online survey, conducted in September 2015, garnered responses from more than 500 executives across a variety of regions, industries, and company sizes.

Box 1. Identifying the most critical internal barriers for organizations (continued)

Setting the right vision and strategy for data and analytics use was the top-rated hurdle among all participants, with more than 45 percent listing it as one of their top three concerns. This is an issue, since personal buy-in from senior management has a direct impact on internal projects. Moving to a model of more data-driven decision making is not as simple as buying a new IT system; it requires leadership to bring about lasting organizational change and usher in a new way of doing business. Executives who reported that their company has made effective use of analytics most often ranked senior management involvement as the factor that has contributed most to their success.

The talent needed to execute the leadership vision is in high demand. In fact, approximately half of executives across geographies and industries reported greater difficulty recruiting analytical talent than any other kind of talent. Forty percent say retention is also an issue. "Business translators" who can bridge between analytics and other functions were reported to be the most difficult to find, followed by data scientists and engineers. Talent scarcity is a major concern that we discuss in greater detail later in this chapter.

A company's IT infrastructure is the backbone through which it can access its data, integrate new data, and perform relevant analyses. Respondents in location-based data services, the area highlighted in our 2011 research that has made the greatest progress in capturing value, reported relatively low barriers in IT infrastructure, an area where heavy investment has paid off. But firms in other industries have struggled to move on from legacy systems that trap data in silos. Even when their companies have allocated investment dollars to upgrading, many executives worry about their ability to choose the most effective systems for their needs. Finally, business units and functions need support and tailoring to use analytics systems effectively; this was a recurring theme from respondents.

Respondents who said their companies had made ineffective use of analytics noted that their biggest challenge was designing the right organizational structure to support it. This needs to include tracking the business impact and making existing processes flexible enough to respond to new data-driven insights. Firms may be comfortable with using analytics in certain areas, but for many, those changes have not filtered through the entire organization.

Talent remains a critical constraint

Human capital has proven to be one of the biggest barriers standing in the way of realizing the full potential of data and analytics. There are four broad types of roles to consider: the data architects who design data systems and related processes; the data engineers who scale data solutions and build products; the data scientists who analyze data with increasingly sophisticated techniques to develop insights; and "business translators" who have both technical and domain- or function-specific business knowledge, enabling them to turn analytical insights into profit and loss impact. In addition to these four categories, data visualization is an important skill set, vital to the last-mile challenge of discovering value. It may be performed by data scientists and business translators, or it can be a stand-alone role—and it is particularly powerful when combined with creative visual design skills as well as experience in creating effective user interfaces and user experiences.

Our 2011 report hypothesized that the demand for data scientists in the United States alone could far exceed the availability of workers with these valuable skills.[40] Since then, the labor market has borne out this hypothesis. As a result of soaring demand for data scientists, their average wages rose by approximately 16 percent per year from 2012 to 2014.[41] This far outstrips the less than 2 percent increase in the nominal average salary across all occupations in Bureau of Labor Statistics data. Top performers with a very scarce skill set, such as deep learning, can command very high salaries. Glassdoor.com lists "data scientist" as the best job in 2016 based on number of job openings, salary, and career opportunities.[42] LinkedIn reports that the ability to do statistical analysis and data mining is one of the most sought-after skills of 2016.[43]

Roles for data scientists are becoming more specialized. On one end of the spectrum are data scientists who research and advance the most cutting-edge algorithms themselves—and this elite group likely numbers fewer than 1,000 people globally. At the other end are data scientists working closer to business uses and developing more practical firm-specific insights and applications.

The scarcity of elite data scientists has even become a factor in some acquisitions. Google, for example, acquired DeepMind Technologies in 2014, at an estimated price of $500 million. With approximately 75 DeepMind employees at the time of the deal, the price tag was nearly $7 million per employee.[44] This is in line with other estimates by experts, who say that "aqui-hires" of cutting-edge AI startups cost around $5 million to $10 million per employee. In this case, the DeepMind acquisition resulted in the development of AlphaGo, which became the first AI program to defeat a human professional player in the game of Go.[45] It also reportedly enabled Google to reduce the cooling costs for its vast data centers by 40 percent, saving several hundred million dollars per year.[46] The DeepMind acquisition could pay off for Google from just this one application alone.

The supply side has been responding to the growing demand for analytics talent. In the United States, students are flocking to programs emphasizing data and analytics. The number of graduates with degrees of all levels in these fields grew by 7.5 percent from 2010 to 2015, compared with 2.4 percent growth in all other areas of study. Universities are also

---

40 In our 2011 analysis, this role was referred to as "deep analytical talent."

41 *Beyond the talent shortage: How tech candidates search for jobs*, Indeed.com, September 2015.

42 "25 best jobs in America," Glassdoor.com blog, available at https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

43 "The 25 skills that can get you hired in 2016," LinkedIn official blog, January 2016, available at https://blog.linkedin.com/2016/01/12/the-25-skills-that-can-get-you-hired-in-2016.

44 Catherine Shu, "Google acquires artificial intelligence startup DeepMind for more than $500 million," TechCrunch, January 26, 2014.

45 See, for example, Christof Koch, "How the computer beat the Go master," *Scientific American*, March 19, 2016. This achievement was widely regarded as a seminal moment in advancing artificial intelligence.

46 See DeepMind corporate blog at https://deepmind.com/applied/deepmind-for-google/.

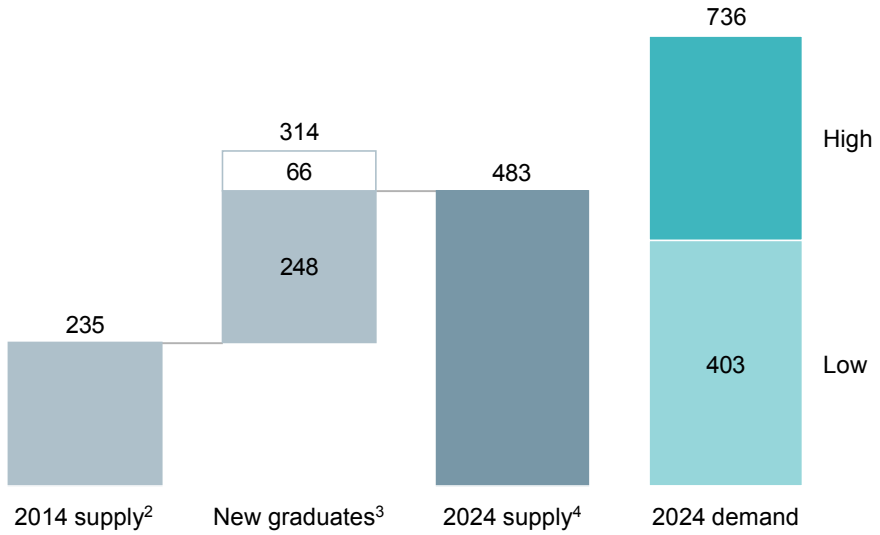
launching programs specifically targeted to data analytics or data science, and business analytics. Currently there are more than 120 master's programs for data analytics or science and more than 100 for business analytics. In addition to formal education tracks, other avenues have opened up to help more people acquire data science skills, including boot camps, MOOCs (massive open online courses), and certificates. However, the scalability and hiring success of these alternative models for data science training remains to be seen. Because data science is highly specialized, the jury is still out on whether employers are willing to hire from non-traditional sources.

In the short run, however, even this robust growth in supply is likely to leave some companies scrambling. It would be insufficient to meet the 12 percent annual growth in demand that could result in the most aggressive case that we modeled (Exhibit 4). This scenario would produce a shortfall of roughly 250,000 data scientists. As a result, we expect to see salaries for data scientists continue to grow. However, one trend could mitigate demand in the medium term: the possibility that some part of the activities performed by data scientists may become automated. More than 50 percent of the average data scientist's work is data preparation, including cleaning and structuring data. As data tools improve, they could perform a significant portion of these activities, potentially helping to ease the demand for data scientists within ten years.

Exhibit 4

The expected number of trained data scientists would not be sufficient to meet demand in a high-case scenario

Supply and demand of data scientists in the United States[1]
Thousand



1 The calculation is across all Standard Occupational Classifications except the ones that are clear false positive hits.
2 2014 fraction per occupation times the 2014 BLS EP employment per occupation, assuming the job market is in equilibrium and supply equals demand. Including 355 occupations.
3 Graduates from US universities during ten years who are estimated to have the skill set required to be data scientists. Includes removing the retirement of 2.1% from current supply.
4 For each industry we calculate the share individually for the top five US companies based on their market capitalization. Professional services is an exception; there we use consulting companies irrespective of their home country.
NOTE: Numbers may not sum due to rounding.

SOURCE: US Bureau of Labor Statistics; Burning Glass; McKinsey Global Institute analysis

On a broader scale, multiple initiatives at the state, national, and international levels aim to develop analytical talent. Examples include the Open Data Institute and the Alan Turing Institute in the United Kingdom; the latter functions also as an incubator for data-driven startups. The European Commission launched a big data strategy in 2014. The United

States published its own federal big data R&D strategic plan in 2016, focusing on the creation of a big data ecosystem that features collaboration among government agencies, universities, companies, and non-profits.

Many organizations focus on the need for data scientists, assuming their presence alone constitutes an analytics transformation. But another equally vital role is that of the business translator who can serve as the link between analytical talent and practical applications to business questions. In some ways, this role determines where the investment ultimately pays off since it is focused on converting analytics into insights and actionable steps. In addition to being data savvy, business translators need to have deep organizational knowledge and industry or functional expertise. This enables them to ask the data science team the right questions and to derive the right insights from their findings. It may be possible to outsource analytics activities, but business translator roles need to be deeply embedded into the organization since they require proprietary knowledge. Many organizations are building these capabilities from within.

2M-4M
projected demand for business translators over the next decade

The ratio of business translators to data scientists needed in a given company depends heavily on how the organization is set up and the number and complexity of the uses the company envisions. But averaging across various contexts, we estimate there will be demand for approximately two million to four million business translators over the next decade. Given that about 9.5 million STEM and business graduates are expected in the United States over this period, approximately 20 to 40 percent of these graduates would need to go into business translator roles to meet demand (though people from other fields can also become business translators). That seems quite aspirational given that today, only some 10 percent of STEM/business graduates go into business translator roles. Two trends could bring supply in line with potential future demand: wages for business translators may have to increase, or more companies will need to implement their own training programs. Some are already doing so, since this role requires a combination of skill sets that is extremely difficult for most companies to find in external hires.[47]

Visualization is an important step in turning data into insight and value, and we estimate that demand for this skill has grown roughly 50 percent annually from 2010 to 2015.[48] Since this is a fairly new development, demand does not always manifest in a specific role. In many instances today, organizations are seeking data scientist or business translator candidates who can also execute data visualization. But we expect that medium-size and large organizations, as well as analytics service providers, will increasingly create specialized positions.

Three trends are driving demand for data visualization skills. First, as data becomes increasingly complex, distilling it is all the more critical to help make the results of data analyses digestible for decision makers. Second, real-time and near-real-time data are becoming more prevalent, and organizations and teams need dynamic dashboards rather than reports. Third, data is increasingly required for decision making through all parts of an organization, and good visualization supports that goal, bringing the information to life in a way that can be understood by those who are new to analytics. New software enables users to make clear and intuitive visualizations from simpler data. But more complex dashboards and data-driven products can require specialized designers. Those who combine a strong understanding of data with user interface/user experience and graphic design skills can play a valuable role in most organizations.

[47] Sam Ransbotham, David Kiron, and Pamela Kirk Prentice, "The talent dividend: Analytics talent is driving competitive advantage at data-oriented companies," *MIT Sloan Management Review*, April 25, 2015.

[48] Based on using the Burning Glass job postings database to search for postings including any of the following skills: data visualization, Tableau, Qlikview, and Spotfire. Normalized with the total number of job postings.

Access to data has improved, but sharing is not always seamless

Overall, industries are less restricted in their ability to access meaningful data than ever before. Data aggregators have found ways to monetize information, governments have opened public data sets and created open data frameworks, and new data sources have continued to proliferate.[49] Firms such as Truven and Explorys are monetizing medical claims data, utilities are sharing detailed information on energy consumption, and sensors embedded in the urban environment are generating valuable data to help manage traffic and complex infrastructure systems. Meanwhile, data policies and regulations have evolved, but barriers remain in some highly regulated sectors, especially health care and the public sector.

Lack of interoperability is a major problem in some industries—notably in health care, where patient information is not always accessible across different EMR systems, and in manufacturing, where various plants and suppliers cannot always seamlessly share information. This is a critical enabler: previous MGI research found that 40 percent of all the potential value associated with the internet of things required interoperability among IoT systems, for example.[50]

In some cases, there are still disincentives to share data. In health care, for example, providers and pharmaceutical companies could stand to lose from greater data sharing with payers. Perhaps the largest hurdle in data access is the need for new data sources to rapidly demonstrate their profit-making potential. Many new data sets are being created in personal health, such as those captured by wearable sensors, but these data sets have yet to demonstrate clinical utility. Given industry dynamics and reimbursement policies, they may experience slow usage and uptake.

Three major concerns continue to challenge the private sector as well as policy makers: privacy, cybersecurity, and liability. All of these can discourage the use of analytics. Privacy issues have been front and center in the European Union, where "right to be forgotten" legislation has required internet companies to take extra steps to clean their records, and citizens have a constitutional right to access data about themselves, even when held by private companies. Meanwhile, privacy concerns have been heightened by repeated cybersecurity breaches. Widely publicized breaches have had major ramifications for companies' relationships with their customers. Many people remain wary about "big brother"–style surveillance by both companies and governments. Customers have reacted negatively to retailers tracking their movements in stores, for example.[51] And lastly, liability frameworks surrounding the use of data and analytics still need to be clarified. In health care, for example, clinical validation can be a lengthy process, and deviating from established guidelines can put physicians or companies at risk of a lawsuit. These concerns will only grow as complicated algorithms play a larger role in decision making, from autonomous driving to deciding where law enforcement resources should be deployed.

•••

Beyond the impact within individual sectors, the soaring demand for data and analytics services has created complex ecosystems. Data may take a long and complex journey from their initial collection to their ultimate business use—and many players are finding ways to monetize data and add value at points along the way. The next chapter examines these ecosystems in greater detail to identify some of these opportunities in this rapidly evolving landscape.

---

[49] *Open data: Unlocking innovation and performance with liquid information,* McKinsey Global Institute, October 2013.

[50] *The internet of things: Mapping the value beyond the hype,* McKinsey Global Institute, June 2015.

[51] Stephanie Clifford and Quentin Hardy, "Attention, shoppers: Store is tracking your cell," *The New York Times*, July 14, 2013.

© Dave and Les Jacobs/Getty Images

3. MAPPING VALUE IN DATA ECOSYSTEMS

Proliferating data sources, including sensors and social media platforms, are creating torrents of information. In fact, so much of the world is instrumented that it is difficult to actually *avoid* generating data. Operating physical systems, processing transactions, sending communications, and movement in a world full of sensors all generate data as a by-product, also known as "data exhaust."

But how much is all this data worth? Data's value comes down to how unique it is and how it will be used, and by whom. Understanding the value in all these small bits of information that need to be gathered, sifted, and analyzed is a tricky proposition, particularly since organizations cannot nail down the value of data until they are able to clearly specify its uses, either immediate or potential. The data may yield nothing, or it may yield the key to launching a new product line or making a scientific breakthrough. It might affect only a small percentage of a company's revenue today, but it could be a key driver of growth in the future.

Many organizations see this potential and are hungry to use data to grow and improve performance—and multiple players are seizing the market opportunities created by this explosion in demand. There are many steps between raw data and actual application of data-derived insights, and there are openings to monetize and add value at many points along the way. As a result, complex data ecosystems have been rapidly evolving.

The biggest opportunities within these ecosystems are unlikely to be in data generation alone, since raw or slightly processed data are usually many steps away from their ultimate, practical use. Furthermore, as data become easier to collect and as storage costs go down, many types of data will become increasingly commoditized, except in certain contexts, where supply is constrained or the data are uniquely suited for high-value uses.

Aggregating information from different sources is critical when a large volume of data is needed or when combining complementary data can lead to new insights. But new tools are allowing end-users to perform this function themselves. Over the longer term, we believe aggregation services will become more valuable only in cases where there are significant barriers to combining data from multiple sources—for example, if a truly massive volume of data is required, if aggregation poses major technical challenges, or if an independent third party is required for coordination.

The most lucrative niches for the future appear to be based in analytics. Previous MGI research has noted that profits are shifting to idea- and knowledge-intensive arenas, including analytical software and algorithms.[52] While companies are often uncertain about what to do with raw data, they are willing to pay for insights that are more readily applicable to strategy, sales, or operations. As organizations become more sophisticated, they are likely to continue devising new ways to collect data—and the demand for insights and analytics will only increase. Since analytics is most effective when combined with deep industry and functional expertise, we expect to see a growing number of specialized players. Furthermore, as the volume of data continues to grow, the ability to separate real insights from the noise will be a source of value. With so much demand, and a scarcity of talent likely to persist in the medium term, firms from throughout the ecosystem are scrambling to take a piece of the analysis market.

---

[52] *Playing to win: The new global competition for corporate profits*, McKinsey Global Institute, September 2015.

NOT ALL DATA ARE CREATED EQUAL—AND THEIR VALUE DEPENDS ON THEIR UNIQUENESS AND END USES

Data has clearly become an important corporate asset—and business leaders want to know how to measure and value the information they hold. Many are asking how much they could earn by selling data to others. But the value of data depends on how they will be used and who will use them. The same piece of data could have different value for different users, depending on their respective economics.

Data have several characteristics that make them unique as an asset. The first is their non-rivalrous nature—that is, the same piece of data can be used by more than one party simultaneously. This makes data similar to other intangible assets such as knowledge and intellectual property. But few organizations list data or information assets on their balance sheets.[53] Most data is monetized indirectly (for example, through selling analytics as a service based on data) rather than through direct sale, which makes its value difficult to disaggregate from other elements in an offering. In some cases, data can be used for barter, which also makes the underlying price tricky to calculate. For example, when a customer signs up for a loyalty card to get discounts or when an individual uses Facebook or Google for free, the customer, wittingly or not, trades personal data for services.

Another important characteristic of data is its sheer diversity. Some of the broad categories include behavioral data (capturing actions in digital and physical environments), transactional data (records of business dealings), ambient or environmental data (conditions in the physical world monitored and captured by sensors), reference material or knowledge (news stories, textbooks, reference works, literature, and the like), and public records. Some data are structured, while images, audio, and video are unstructured. Data can come from a diversity of sources, such as the web, social media, industrial sensors, payment systems, cameras, wearable devices, and human entry. Each of these features relates to how the data are generated and therefore affect their supply.

On the demand side, the value of data depends on how the data will (or could) ultimately be used. Sometimes the value of a piece of data is known because it is directed to a specific, clear purpose—such as when an advertiser purchases information on TV ratings, or when a political campaign purchases an email list from an advocacy group to reach potential voters. An online retailer can measure the difference in conversion from a generic Facebook ad versus one that is based on the customer's browsing history on the retailer's website. In other cases, organizations are unsure about how and where data could be put to work. For instance, vast volumes of data exhaust are generated by 30,000 sensors on a modern oil rig, but less than 1 percent is used in decision making.[54] But one organization's data exhaust could be another organization's data gold. Only by understanding the potential uses of data by individual players can the value of data be determined.

[53] In the case of valuation in merger or acquisition, the value of the target's data assets may partially be recorded as "databases" under intangible assets, and part of the value is captured as goodwill.

[54] *The Internet of Things: Mapping the value beyond the hype*, McKinsey Global Institute, June 2015.

There are many types of uses for data; several examples are described below.

- **Internal cost and revenue optimization:** The potential applications here are numerous. On the cost side, data can be put to work in predictive maintenance, talent and process management, procurement, and supply chain and logistics planning. On the revenue side, insights from data can be used to enter new markets, micro-target customer segments, improve product features, and make distribution channels more effective. Data derived from machines and processes, especially from IoT sensors and from customer behavior and transactions, are most useful for optimizing operations. While the data may be generated internally, there can be opportunities for service providers, since analyzing data is not yet a core competency for many firms in traditional industries. But other companies will build the internal capability to take advantage of their own data to improve their operations.

- **Marketing and advertising:** These functions generally rely on customer transactional and behavioral data aggregated from multiple sources such as social media profiles, demographic information, online browsing history, and previous purchases. It may take a large volume of data to yield insight, but it is critical to have the capability to cut through the noise with effective analysis. Advertising technology firms and social media networks with large amounts of consumer behavioral data are some of the fastest-growing players in this space. There is a direct link from this data to value (as measured through conversion).

- **Market intelligence:** Market intelligence data is compiled with an economy-wide, regional, industry-specific, functional, or market perspective to deliver strategic insights. Many traditional providers of information services fall into this category. A few firms such as Google can generate macrodata themselves, but most providers aggregate from external sources. Because the data provides a clear value to customers and is not easily replicated, it can be sold directly, and its value is easier to determine.

- **Market-making:** Market-making firms, from ride-sharing apps to dating sites, play the role of matching the needs of buyers and sellers. These firms often create platforms to collect the necessary information to enable efficient and effective matching. In some cases, pure signaling data is all that is needed. But in other cases, preference data, reputation data (to validate the authenticity and quality of participants), and transaction and behavior data are crucial. Scale and network effects are important here, as buyers and sellers demand a marketplace that is liquid enough to match their needs efficiently.

- **Training data for artificial intelligence:** Machine learning and deep learning require huge quantities of training data. Some is generated through repeated simulations (such as game playing), some is generated in the public sphere (such as mapping and weather data), and other data is aggregated from a diversity of sources (such as images and video or customer behavior). Public sources and private aggregators can play a crucial role here. Firms that produce huge quantities of relevant data with their own platform have a head start—and first movers may have an advantage, as their offerings will have more time to learn and generate additional data, fueling the virtuous cycle. However, because there is a great variety of potential uses for different users, valuing data here can be challenging.

These ecosystems may overlap. In some cases, the same piece of data (such as customer behavior data) can have multiple applications, each with a different value.

INFORMATION SOMETIMES TAKES A COMPLICATED PATH THROUGH A DATA ECOSYSTEM

Raw data goes through multiple steps before it is actually used. It is aggregated and enhanced when blended with other data sources, and this enhanced data is ultimately transformed into results or insights by applying analytical techniques. Data can be directly or indirectly monetized at various points along the way. In general, the more data is refined, the more it becomes applicable to specific uses, driving up its value. To better understand these dynamics, we examine various roles within data ecosystems, recognizing that some players might fill more than one role.

- **Data generation and collection:** The source and platform where data are initially captured.
- **Data aggregation:** Process and platforms for combining data from multiple sources.
- **Data analysis:** The gleaning of insights from data that can be acted upon.
- **Data infrastructure:** The hardware and software associated with data management. We recognize that a diverse landscape of infrastructure providers offers the hardware and software necessary to execute on the collection, aggregation, and analysis activities described, but that portion of the ecosystem is not the focus of our discussion.

To give one example of how different types of data move through each part of an ecosystem, consider what happens when a consumer applies for a credit card. Consumers generate data when they use and make payments on their existing financial products, including mortgages, credit cards, and other lending accounts. The various financial institutions that hold these accounts collect, organize, and summarize this information generated by the consumer's behavior as a borrower. These entities then share these summary data with credit bureaus, which play the aggregator role by combining them with data from other entities that report late payments (such as collection agencies for medical or utility bills) and from public records such as bankruptcy filings. The credit bureaus are then able to form a more complete view of the customer's credit behavior, and they apply analytics to generate a credit score. A financial institution considering a consumer's application for a card will pay credit bureaus for access to the score and the full credit report to inform its own decision-making models and determine whether to issue the new card.

In this case, the generator of the data (the consumer) does not necessarily own the data. The consumer's agreements with various lenders (which are shaped by legal frameworks) outline how this information can be shared. The process is also complicated by the fact that analysis occurs at different points in the process. Monetization similarly happens at two points: when the credit bureau sells the credit report, and from fees after the bank issues a new credit card.

Within these ecosystems, we believe that value will tend to accrue to providers of analytics (Exhibit 5). It seems likely to shift away from data collectors and aggregators unless particular conditions make those roles more valuable (for example, when data are difficult to collect, proxies are limited, or certain types of data are uniquely necessary for high-value applications). In the pages that follow, we will look at the generation and collection, aggregation, and analysis components of the ecosystem.

Exhibit 5

Within the data ecosystem, value will tend to accrue to firms doing analysis and only under certain conditions to those providing data generation or aggregation

	Data generation and collection	Data aggregation	Data analysis
Description	▪ The source and platform where data is initially captured	▪ Combining data from multiple sources ▪ Forming marketplaces for commercial exchange of data	▪ Manipulating aggregated data to develop actionable insights
Factors driving value up	▪ Certain data types will have higher value if collection barriers are extremely high or data cannot be legally shared between parties	▪ Demand growth as more applications are developed ▪ Value will be higher if aggregation is technically challenging or requires a neutral third party	▪ Talent shortage ▪ Deep sector expertise needed to deliver effectively ▪ Close relationship to actual use or implementation clarifies value
Factors driving value down	▪ Growth in available proxies and expansion of open access will increase supply	▪ Technology advances making aggregation easier	▪ Scope could be limited as solutions will be for "vertical" applications
Future trajectory of value			

SOURCE: McKinsey Global Institute analysis

DATA GENERATION AND COLLECTION IS THE STARTING POINT

The value in data collection, as in any other market, is driven by supply and demand forces. As the supply of data available from new sources continues to expand, we expect that the generation of specific types of raw data will generally become a less valuable function over time. The exceptions to this will be certain cases in which supply is constrained or demand suddenly spikes for certain types of data.

On the supply side, the market for any particular type of data is shaped by difficulty of collection, access, and availability of substitute data (proxies from which similar insights can be drawn). The volume of available data has skyrocketed, and this increase in supply drives down the relative value of any one piece of information. More data is available than ever before as people spend more time online, as sensors are deployed throughout the physical environment, and as formerly manual processes (such as toll collection) are automated. Some kinds of data that were labor-intensive to collect in the past, such as TV watching habits, are now collected in the normal course of business.

This trend shows no sign of slowing. The norm of digital media, social networking, and search businesses is to give content away for free and "monetize the eyeballs" in the form of advertising, generating user data in the process. The internet amplified a culture of openness that had already taken root in the tech world, and non-profits (such as Wikimedia and the Open Data Institute) have emerged to support open data initiatives. The public sector, always a large-scale data collector, has invested in making its data more easily available. City governments, for instance, have made significant amounts of information public, from restaurant health inspection scores to school performance ratings and crime statistics. Previous MGI research estimated that open data can help unlock $3 trillion to $5 trillion in economic value annually across seven sectors.[55]

[55] *Open data: Unlocking innovation and performance with liquid information*, McKinsey Global Institute, October 2013.

Together, the growth in data generation and the expanding accessibility of data are increasing the availability of proxies. Today users are more likely to find a free or low-cost substitute for data that were once scarce. Google Trends, for instance, offers a free proxy for public sentiment data that previously would have been collected through phone surveys. As proxies improve and proliferate over time, the value of any single piece of data will fall.

On the demand side, the market is shaped by ease of use, network effects, and value of the ultimate uses of data. Thanks to improvements in data's ease of use, there are simply more users consuming data than ever before. Once only trained data scientists could perform certain analyses, but now more user-friendly tools make it possible for business analysts to use large data sets in their work. At the same time, machine learning advances have expanded uses for unstructured data such as images, driving up the demand for these types of data. For example, Lose It, a calorie-tracking app, allows users to log their intake by taking a picture of their food and matching it with an image database.[56] The more users upload and validate that their image matches the suggested item, the more effective the algorithm grows. This kind of dynamic establishes a virtuous cycle of demand for these forms of data.

Another factor driving demand is that different types of data are being linked to derive new insights. A mining company, for example, can use weather data to identify lightning strikes as a proxy for copper deposits. Linking data leads to more applications and more demand for data, and we expect the trend to grow as organizations become more sophisticated in their data and analytics capabilities.

Finally, demand for data is driven by the expected value associated with their ultimate use. In many cases, organizations are still learning what to do with all the data that is suddenly available. But as analytics tools improve, so do the odds that at least some value can be extracted from any given data set.

On balance, the supply and demand factors described above suggest data will continue to become commoditized. Certain types of data that were expensive to obtain in previous years are now free, such as $1,400 encyclopedia sets replaced by Wikipedia. Instead of dialing 411 and paying $1 per query to find a phone number, smartphone users can use $0.005 worth of data on their device to search for and call any number they need.[57] The non-rivalrous nature of data (that is, the fact that it can be used by different parties simultaneously) may contribute to making it more of a commodity.

But there are some exceptions to the commoditization trend. When access is limited by physical barriers or collection is expensive, data will hold its value. Google, for example, made huge investments to collect data for Google Maps, driving more than seven million miles to gather relevant images and location data.[58] These data sources are public, but the barriers to collection at this scale remain extremely high. The holders of these kinds of unique, hard-to-capture data sets will continue to have opportunities to directly monetize data. In these cases, data's non-rivalrous nature will help the provider of this data to capture more value, since there are opportunities to monetize the data more than once. Some firms may hold proprietary data that has staying power as the "industry standard" (as with Nielsen or the College Board). It is also possible to artificially create scarcity for certain types of data through legal means, such as licensing to prevent its transfer to other users.

[56] Lora Kolodny, "Lose It launches Snap It to let users count calories in food photos," TechCrunch, September 29, 2016.

[57] $0.005 assumes 0.4 megabytes used to load search page and a monthly data plan of $12.50 per gigabyte.

[58] Greg Miller, "The huge, unseen operation behind the accuracy of Google Maps," *Wired*, December 8, 2014.

Data collected in private settings can also resist commoditization. It is difficult to imagine a proxy that could easily substitute for the sensor data collected in a jet engine for predictive maintenance purposes. Raw data produced by auto telemetry for insurance purposes are quite valuable to automakers and others, and the company that collects the data has opportunities to monetize that data accordingly. However, since there are often controls over personally identifiable data, owners of this kind of proprietary data are constrained in what they can sell externally. Therefore the value these data collectors can capture is usually limited to indirect monetization in the form of internal improvements or value-added services.

An important case of value accruing to firms that engage in data generation and collection involves market-making platforms with network effects. Since suppliers will want to go where demand is, and consumers will want to go where suppliers are, these platforms have natural network effects. On these platforms, data beget more data. Social media platforms also have significant network effects since individuals naturally want to be on platforms where their friends and others are also present. Search platforms generate data when users search, which enables their search algorithms to produce higher-quality results, which draws more users to using the search engine, and on and on. The network effects of such platforms often lead to a small number of players collecting and owning a large percentage of data generated in these ecosystems. Proxies may also be sparse as the most valuable data will be specific platform behavior data that are unique to the users on the platform. In these circumstances, a few platform leaders will be able to capture significant value from owning their user behavior data. These network effects can also arise in the aggregation part of the value chain, which we discuss further below.

In the absence of these types of exceptional supply constraints, simply selling raw data is likely to generate diminishing returns over time. But in situations where these constraints exist (or where a business model creates them), generators and collectors of data can capture significant value.

DATA AGGREGATION COMBINES DATA FROM MULTIPLE SOURCES TO CREATE A MORE COMPLETE PICTURE

Firms aggregating data from different sources can capture value for a few reasons. First, they can serve as a "one-stop shop" for data from multiple sources. Second, data aggregation adds value because combined data may yield better insights. Benchmarking the performance of multiple entities may help identify areas for improvement, for example. Risk assessments are more accurate when they incorporate multiple pieces of behavioral and environmental data. Combining mobile and desktop browsing behavior offers a more complete picture of a customer's consumption patterns. Going back to our credit example, a consumer's history of timely mortgage payments is more valuable when joined with data on how that same consumer handles credit cards. Third, third-party aggregation can be useful in circumstances where there is a collective action problem among competitors.

Aggregation can produce significant value, but it is becoming easier for users to perform many aspects of this function themselves. There has been robust growth in new software services for organizing data from different internal and external sources; this niche has attracted significant venture capital. End-users now have cheaper and more powerful tools to aggregate data on their own.

The value of aggregation therefore seems likely to increase only in cases where integrating data from various sources is challenging. In these cases, aggregators may find that the information they synthesize can be sold to a broad range of customers. Location data, for example, is huge and noisy, and data that are validated from multiple sources may be more valuable to customers across industries that want to deliver targeted advertising and offers. Aggregating location data, as well as other very large data sets such as those taken from sensors, will be an increasingly valuable function.

Aggregation services are particularly valuable when combining and processing data are technically difficult, or when coordinating access across diverse sources is a barrier. This can be a complex process even if the underlying data are commoditized (as with financial market data) or when they are highly varied and differentiated (as with health records). Many traditional marketing data providers (such mailing list vendors) and information services providers (such as Westlaw, Bloomberg, and Argus) fall into this category and have developed long-standing relationships with data collectors or have technical assets that enable aggregation. Many of these aggregators also serve as "data guides," using their deep understanding of complex data environments and privacy regulations to advise their clients on how to best handle the data.

Financial data aggregators capture value by resolving the technical challenges of gathering and distilling huge amounts of information. Much of the data they combine are publicly available from public filings and financial markets. But the data come in massive volumes, and real-time updates are critical for investment decisions. Financial data aggregators such as Bloomberg and Capital IQ have developed software to automate data aggregation. Like data collectors, however, they face a risk that barriers to entry may fall as technology improves, allowing new firms to offer similar services at a cheaper price. But users' familiarity with the incumbent services, along with trust in these providers as industry standards, can help to preserve value capture by these players.[59] In some cases, companies can build a business model around serving as a third-party aggregator for competitors within a given industry. Argus, for example, aggregates data from across financial institutions to provide benchmarks. Data are collected by the participating financial institutions, but those institutions do not share information directly with one another because of competitive concerns. As an aggregator, Argus resolves the coordination challenge by serving as an intermediary between these institutions. Each one submits standardized data, and Argus summarizes the inputs as industry averages. Each institution then receives a report showing its own submission compared with the group average, allowing it to gauge its performance across the industry as a whole.

The third-party aggregator model can be relevant in any industry with fragmented competitors (such as the travel ecosystem, for example). This model has the potential to create network effects, as individual organizations have incentives to share their data and join an aggregation service that already has many members and lots of data. The more members and data in any aggregation service, the more benefit each additional member gets from joining. This kind of effect often leads to just one aggregator or a small number of aggregators dominating any given ecosystem.

Another emerging model in aggregation involves data marketplaces that match parties that want to purchase data with those that can supply it. They can fill a valuable role in cases where data collectors and end-users are highly fragmented. Marketplaces focus on creating the technical platform and quality standards needed for a broad range of firms to participate. They typically aggregate data but do not integrate them; that task is left to the end-users. The presence of marketplaces lends greater liquidity to the exchange of data,

[59] Providers of information services in finance can generate operating margins of more than 25 percent.

often facilitating commercial arrangements between unrelated parties that previously would not have been able to monetize their data.

DATA ANALYSIS IS WHERE VALUE WILL ACCRUE IN MOST ECOSYSTEMS

Translating data into business insights is generally the most important—and most valuable—step in the data ecosystem. This is where data meet their use and user. Great analysis can squeeze insights even from mediocre data, while conversely, bad analysis can destroy the potential value of high-quality data.

On the demand side, since analysis is often the last step, the value generated by data and analysis becomes much clearer. This puts the analytics provider in a better position to capture a portion of this value. While companies are often uncertain about what to do with a huge volume of newly collected raw data, they are willing to pay for insights that are directly related to sales, strategy, and other business functions. Across sectors, firms have a larger appetite for data as improved analytical tools and capabilities open up new uses. On the supply side, the highly specialized talent needed for analytics and interpretation is scarce. Providing data analytics requires industry and functional expertise, and there is a limited pool of talent and organizations combining these skill sets. Even as tools and platforms improve, the need to combine analytical and domain expertise will continue to present a bottleneck, driving up the value of analytics.

Third-party analytics and interpretation offerings can be delivered as software, a customized service, or a product. The makers of analysis software and tools have some of the highest net margins in the data ecosystem, often in the range of 20 to 30 percent.[60] But purchasing the tools alone may not be enough for companies that are short on technical talent or need customized solutions, and many are turning to external service providers. That being said, the majority of the value generated by analytics will be captured by the user of the analysis.

The most successful analytics providers combine technical capabilities with industry or functional expertise. Performing analytics that support predictive maintenance, for example, requires a deep understanding of how machines work. Using analytics to optimize energy consumption starts with a thorough understanding of the systems that use energy. This type of understanding gives these providers the advantage of knowing just how much value they are creating for their clients, and they can price accordingly. But expertise in one application in one sector may or may not be immediately transferable to another industry or problem type; each vertical problem has to be attacked individually. Unlike data aggregation, which is a horizontal play across different types and sources of data, analytics is a vertical play where each additional vertical use case requires additional domain knowledge and sometimes entirely new analytic techniques.

Given the size of the opportunities in analysis, firms in other parts of the ecosystem often add analytics to claim a piece of this high-value segment of the market. Some data collectors or aggregators are adding new data product offerings. In insurance, for example, CoreLogic has used its data assets to develop catastrophic risk management products—in this case, complex, robust catastrophe risk scores that are sold to insurers. Other data collectors and aggregators are making similar moves into high-value data products, such as Wood Mackenzie in energy and Inovalon in health care.

In addition to the data products mentioned above, data collectors and aggregators are offering to integrate with clients' data and perform analysis as a service, especially in industries such as health care where large-scale analytics is needed but where most firms lack that core competency. Startups such as SparkBeyond are integrating aggregation and analytic capabilities to apply machine learning to specific business challenges, with

[60] Based on analysis of Wikibon Big Data Market Forecast.

capabilities far beyond what most companies would be able to build internally. In some instances, the makers of analytics software will expand into providing analytical services that build on their specific area of expertise. Some marketing agencies as well as consulting and business intelligence firms are adding analytics offerings, sometimes acquiring smaller analytics platforms to boost their capabilities. Firms offering integrated technology solutions are also emphasizing analytics, as IBM has done with IBM Watson.

•••

The rapid evolution of data ecosystems points to the scope of the data analytics revolution that is beginning to take hold. The shift toward data-guided decision making is transforming how companies organize, operate, manage talent, and create value. In short, big data could generate big disruptions. The next chapter looks at six breakthrough models and capabilities that could set off a wave of change across industries and in society.

© Lane Oatey/Blue Jean Images/Getty Images

© RyanJLane/Getty Images

4. MODELS OF DISRUPTION FUELED BY DATA AND ANALYTICS

Data and analytics are already reshaping certain industries—and the ripple effects are poised to grow in the decade ahead, extending into new sectors and many parts of society.

This chapter considers how these changes are playing out. But instead of taking an industry-by-industry view, we believe it is more useful to consider the types of disruptive models that data and analytics enable. This is by no means an exhaustive list, and the lines separating some of these models are not clear-cut, as some companies combine these approaches. Nevertheless, business leaders can apply this lens to the markets in which they operate with an eye toward preparing for what comes next.

Certain characteristics of a given market open the door to data-driven disruption (Exhibit 6). Data collected in one domain can be deployed, for unrelated purposes, in an entirely different industry. Using Google search data to create a price index and drawing on credit scores to inform auto insurance rates are prime examples of how "orthogonal data" can be put to work in solving different types of business problems. Collecting new types of data could set off cascading industry effects.

Exhibit 6

Data and analytics underpin six disruptive models, and certain characteristics make individual domains susceptible

Indicators of potential for disruption:

- Assets are underutilized due to inefficient signaling
- Supply/demand mismatch
- Dependence on large amounts of personalized data
- Data is siloed or fragmented
- Large value in combining data from multiple sources
- R&D is core to the business model
- Decision making is subject to human biases
- Speed of decision making limited by human constraints
- Large value associated with improving accuracy of prediction

Archetype of disruption	Domains that could be disrupted
Business models enabled by orthogonal data	▪ Insurance ▪ Health care ▪ Human capital/talent
Hyperscale, real-time matching	▪ Transportation and logistics ▪ Automotive ▪ Smart cities and infrastructure
Radical personalization	▪ Health care ▪ Retail ▪ Media ▪ Education
Massive data integration capabilities	▪ Banking ▪ Insurance ▪ Public sector ▪ Human capital/talent
Data-driven discovery	▪ Life sciences and pharmaceuticals ▪ Material sciences ▪ Technology
Enhanced decision making	▪ Smart cities ▪ Health care ▪ Insurance ▪ Human capital/talent

SOURCE: McKinsey Global Institute analysis

In some domains, supply and demand are matched inefficiently. This may result in the underutilization of assets, among other problems. Digital platforms that offer large-scale, real-time matching with dynamic pricing could revolutionize markets with these

characteristics—a list that includes areas like transportation, hospitality, certain labor markets, energy, and even some public infrastructure.

In markets where incumbents have traditionally relied on broad demographics or blunt measures to segment customers, new players could change the game by using finer behavioral data to customize products, services, and marketing for individuals. This could have far-reaching impacts in many areas, including health care, media, consumer products and retail, education, and even political campaigns. In industries where integrating more and better data can dramatically improve performance—such as banking, insurance, retail, the public sector, and beyond—the organizations that master this capability can realize major advantages. Data and analytics could similarly shake up industries where the business model is built on innovation and discovery. Machine learning and deep learning could fuel everything from drug development to engineering to user experience and design.

Additionally, any endeavor that can be marred by human error, biases, and fallibility could be transformed by what is perhaps the most profound capability of data and analytics: the ability to enhance, support, and even automate human decision making by drawing on vast amounts of data and evidence. Technology now presents the opportunity to remove human limitations from many situations and to make faster, more accurate, more consistent, and more transparent decisions. This capability has wide applicability for businesses in every industry as well as in many areas of society and daily life.

This chapter describes the new data that is core to each of these models. It shows how that new data is being used to shake up specific parts of the economy—and points to other areas where disruption could occur in the near future.

BRINGING ORTHOGONAL DATA TO BEAR CAN CHANGE THE BASIS OF COMPETITION

Data and analytics have been part of the operating model to at least some degree in multiple industries for years. Insurers offering automobile, property, and casualty insurance, for instance, long ago began incorporating risk factors such as demographics (age, place of residence, years of driving experience, and so forth) into the analytics that underpin their underwriting decisions.

But as discussed in Chapter 3, data are proliferating. Many new types, from new sources, can be brought to bear on any problem. In industries where most incumbents have become used to relying on a certain type of standardized data to make decisions, bringing in fresh types of data sets to supplement those already in use can change the basis of competition. New entrants with privileged access to these types of “orthogonal” data sets can pose a uniquely powerful challenge to incumbents.

Returning to our property and casualty insurance example above, we see this playing out. New companies have entered the marketplace with telematics data that provide insight into driving behavior. This data set is orthogonal to the demographic data that had previously been used for underwriting. Other domains could be fertile ground for bringing in orthogonal data from the internet of things. Health care has traditionally relied on medical histories, medical examinations, and laboratory results, but a new set of orthogonal data is being generated by consumer health devices such as wearables and connected health devices in the home (such as blood pressure monitors or insulin pumps). Some innovators are experimenting to determine if data from these devices, while not clinical grade, could enhance wellness and health. Connected light fixtures, which sense the presence of people in a room and have been sold with the promise of reducing energy usage, generate “data exhaust” that property managers can use to optimize physical space planning in future real estate developments. Even in human resources, some organizations have secured employee buy-in to wear devices that capture data and yield insights into the “real” social

networks that exist in the workplace, enabling these organizations to optimize collaboration through changes in work spaces.

Orthogonal data will rarely replace the data that are already in use in a domain; it is more likely that an organization will integrate the orthogonal data with existing data. In the pages that follow, we will see several examples of orthogonal data being combined with existing data as part of a disruptive business model or capability.

HYPERSCALE PLATFORMS CAN MATCH SUPPLY AND DEMAND IN REAL TIME

Data and analytics are transforming the way markets connect sellers and buyers for many products and services. In some markets, each offering has critical variations, and the buyer prioritizes finding the right fit over the speed of the match. This is the case in real estate, for example, where buyers have strong preferences and finding exactly the right house is the priority. In others, the speed of the match is critical. "Hyperscale" digital platforms can use data and analytics to meet both types of needs.

These platforms have already set off major ripple effects in urban transportation, retail, and other areas. But that could be only the beginning. They could also transform energy markets by enabling smart grids to deliver distributed energy from many small producers. And they could make labor markets more efficient, altering the way employers and workers connect for both traditional jobs and independent work.

These platforms are already transforming the market for transportation

Hyperscale digital platforms can have notable impact in markets where demand and supply fluctuate frequently, where poor signaling mechanisms produce slow matches, or where supply-side assets are underutilized. These characteristics describe the status quo that prevailed in the taxi industry for many years before the arrival of Uber, Lyft, Didi Chuxing, and similar services. Conventional taxicabs relied on crude signaling mechanisms—literally, in this case, a would-be passenger attempting to wave down an empty cab in the street or calling a company's dispatcher. These mechanisms created significant unmet demand. On the supply side, many cabs spent a large share of their time empty and cruising for passengers. Furthermore, most vehicles are underutilized; globally, most personally owned cars are in use for approximately 5 to 10 percent of waking hours.[61] Excess supply sometimes pooled in certain spots, while other areas went largely underserved. For several reasons, including heavy regulation and static pricing, taxi markets were and continue to be highly inefficient. These inefficiencies—combined with the fact that the speed of hailing is of primary importance—made the market ripe for a radically different model to take root.

That model combined digital platforms with location-based mapping technology to instantly match would-be passengers with the driver in closest proximity. In addition, the location data can be analyzed at the aggregate level to monitor overall fluctuations in supply and demand. This allows for dynamic pricing adjustments, with price increases creating incentives for more drivers to work during periods of high demand. The platform nature of these services, which makes it easy for new drivers to join, unleashed flexible supply into the transportation market. Different types of mobility services have been launched, including not only ride sharing (such as Uber and Lyft) but also car sharing (Zipcar) and ride pooling (Lyft Line, UberPool).

From the outset, these platforms collected data from their user base to implement improvements—and as the user base grew, they generated even more data that the operators used to improve their predictive algorithms to offer better service. This feedback mechanism supported exponential growth. Uber, founded in 2009, is now in more than

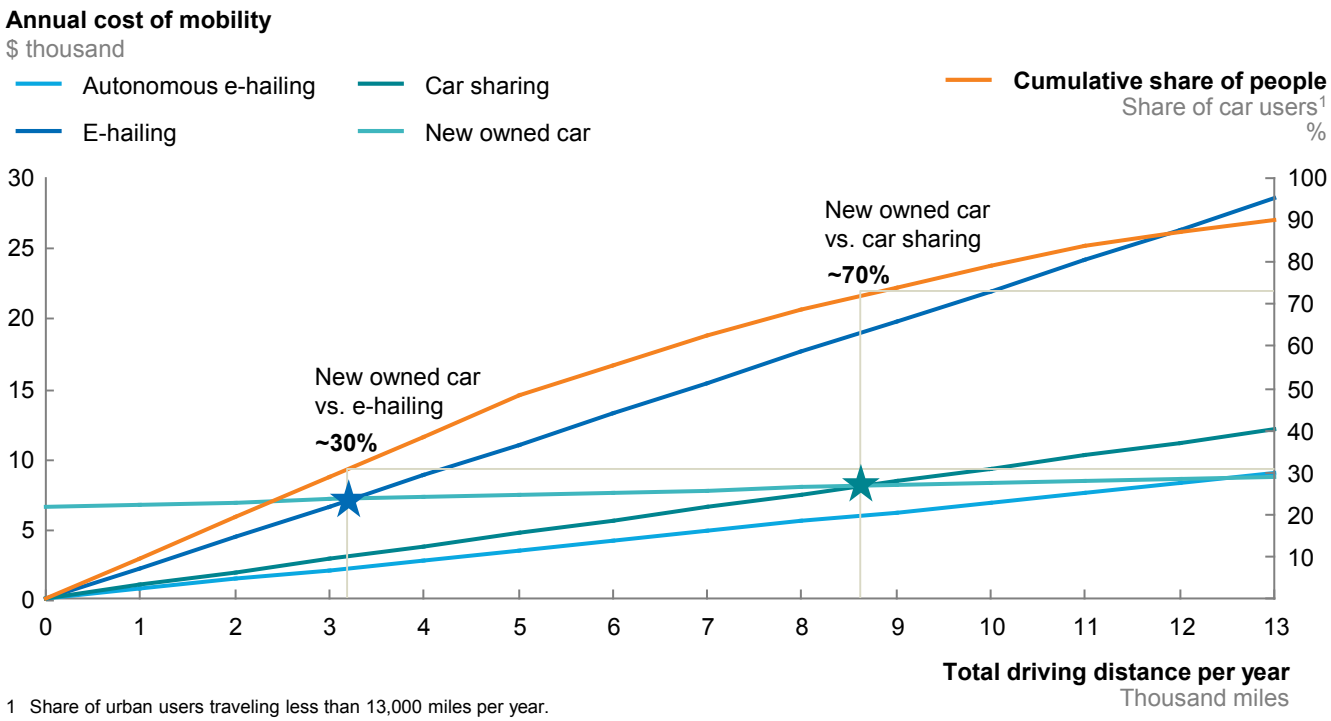
[61] See Paul Barter, "'Cars are parked 95% of the time.' Let's check," ReinventingParking.org, February 22, 2013, for a roundup of studies on the utilization rates of cars, including research from economist Donald Shoup.

500 cities and delivered its two billionth ride in the summer of 2016.[62] Lyft reportedly hit almost 14 million monthly rides in July 2016.[63] In China, ride-sharing giant Didi Chuxing now matches more than ten million rides daily.[64] Today mobility services account for only about 4 percent of total miles traveled by passenger vehicles globally. Based on their growth momentum, this share could rise to more than 15 to 20 percent by 2030. This includes only real-time matching platforms and excludes the potential effects of autonomous vehicles.[65]

The changes taking place in urban transportation—including a substantial hit to the taxi industry—may be only the first stage of an even bigger wave of disruption caused by mobility services. These services are beginning to change the calculus of car ownership, particularly for urban residents. Exhibit 7 indicates that almost one-third of new car buyers living in urban areas in the United States (the segment who travel less than 3,500 miles per year) would come out ahead in their annual transportation costs by forgoing their purchase and relying instead on ride-sharing services. For them, the cost of purchasing, maintaining, and fueling a vehicle is greater than the cost of spending on ride-sharing services as needed. If we compare car ownership to car sharing instead of ride sharing, around 70 percent of potential car buyers could benefit from forgoing their purchase. A future breakthrough that incorporates autonomous vehicles into these services, thereby reducing their operating costs, could increase this share to 90 percent of potential car buyers in urban settings.

Exhibit 7

Mobility services could drastically change the calculus of car ownership



1 Share of urban users traveling less than 13,000 miles per year.

SOURCE: McKinsey Automotive 2030 report, Urban mobility survey; McKinsey Global Institute analysis

[62] Fitz Tepper, "Uber has completed 2 billion rides," TechCrunch, July 18, 2016.

[63] Johana Bhuiyan, "Lyft hit a record of 14 million rides last month with run-rate revenue of as much as $500 million," ReCode, August 2, 2016.

[64] Steven Millward, "China's Didi Chuxing now sees 10 million daily rides," Tech in Asia, March 22, 2016.

[65] "Shared mobility on the road of the future," Morgan Stanley Research, www.morganstanley.com/ideas/car-of-future-is-autonomous-electric-shared-mobility.

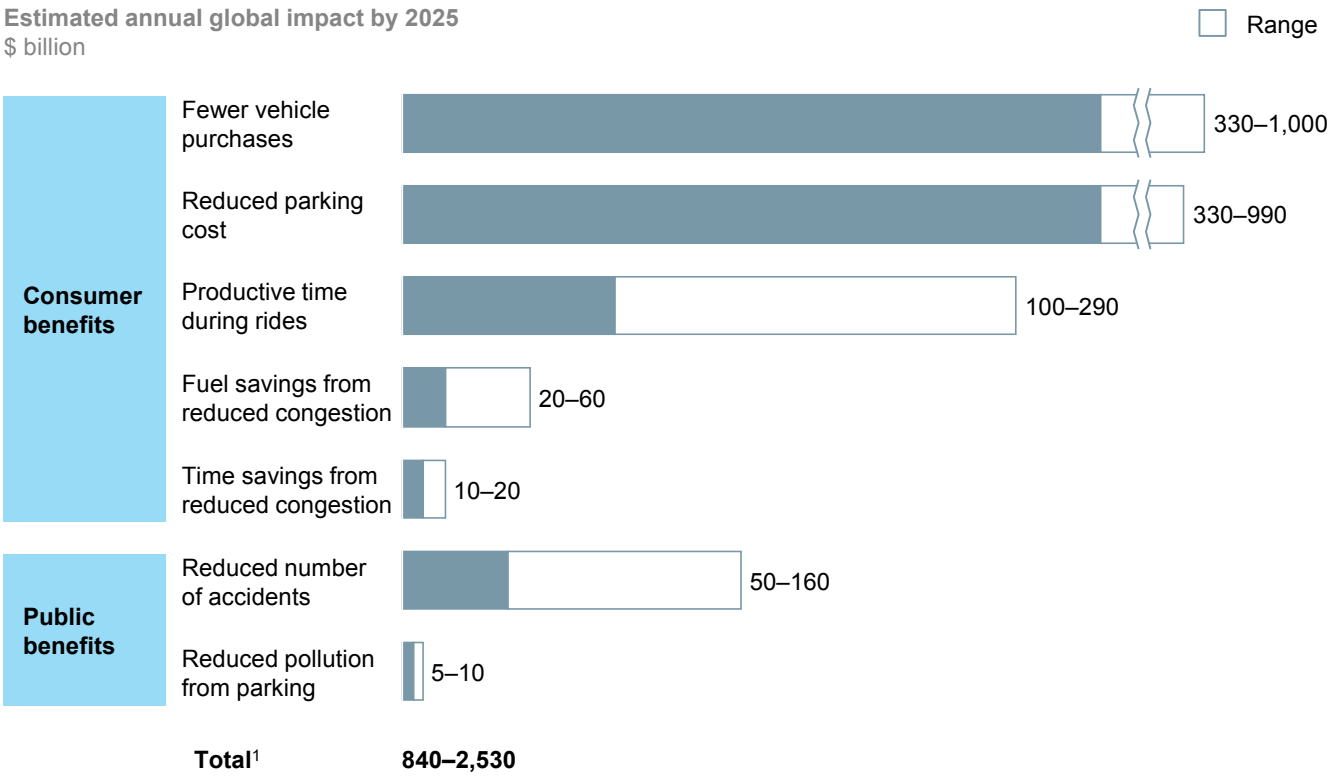
These trends are beginning to reshape the structure of the overall transportation industry. Value is already shifting from physical assets to data, analytics, and platforms as well as high-margin services such as matching. This is even playing out within the car-sharing market itself, as Car2Go, Zipcar, and other firms that own fleets now face newer platform-based players such as Getaround. Hyperscale platforms will likely create concentrated markets, since network effects are crucial to their success.

Economic impact and disruption

Hyperscale, real-time matching in transportation has the potential to generate tremendous economic impact. Individual consumers stand to reap savings on car purchases, fuel, and insurance by shifting to mobility services; they could also gain from having to spend less time looking for parking. Furthermore, the public will benefit from reduced real estate dedicated to parking, improved road safety, and reduced pollution. Summing these effects and assuming a 10 to 30 percent adoption rate of mobility services among low-mileage travelers, we estimate global economic impact in the range of $845 billion to some $2.5 trillion annually by 2025 (Exhibit 8).[66] However, these shifts will create winners and losers. Some of the benefits will surely go to consumer surplus, while some will go to the providers of these platforms and mobility services.

Exhibit 8

Hyperscale real-time matching in transportation could potentially create some $850 billion to $2.5 trillion in economic impact



1 Roughly 60 percent of the economic impact occurs in developed economies, and the remainder in emerging economies.
NOTE: Assumes a 10–30% adoption rate among low-mileage urban travelers. Numbers may not sum due to rounding.

SOURCE: McKinsey Global Institute analysis

[66] These estimates assume 10 to 30 percent adoption of mobility services globally.

The largest part of this impact stems from consumers shifting away from car ownership. By moving to mobility services, the average global consumer is likely to gain around $2,000 annually in cost savings over purchasing and maintaining a vehicle. With around 1.6 billion active vehicles projected by 2025, even 10 to 30 percent adoption of mobility services among low-mileage vehicle users can lead to $330 billion to $1 trillion in savings annually (this includes about $140 billion to $430 billion potentially realized in developed regions and $190 billion to $570 billion in developing regions).

Up to **$1T** potential consumer savings from adopting mobility services rather than purchasing vehicles

In addition to direct car ownership costs, the shift toward mobility services will generate significant savings in related costs like parking. Consumers are projected to spend $3.3 trillion on parking services in 2025, but the use of mobility services could allow them to save $330 billion to $990 billion. Around $220 billion to $650 billion of this could be realized in developed countries and $110 billion to $340 billion in developing countries.

There is an additional benefit from the reduced demand for driving and parking. If 15 to 30 percent of drivers on the road in cities are looking for parking, this is a major logistical challenge in dense urban cores. By boosting the utilization of each vehicle, mobility services can decrease demand for parking and help reduce congestion, which creates further positive ripple effects on mobility and time saved. The reduced search for parking can generate a time-saving effect due to reduced congestion that can be valued at $10 billion to $20 billion as well as fuel savings in the range of an additional $20 billion to $60 billion.

Meanwhile, the shift to mobility services can improve productivity. Each day, workers spend 50 minutes in driving commutes on average in both developed and developing countries.[67] If even half of that time can be used more productively for work, mobility services could generate an additional $100 billion to $290 billion in potential benefit.

Finally, ride sharing can improve road safety by creating a more viable option that keeps people from getting behind the wheel when they have been drinking, they are excessively tired, or they have other impairments (such as difficulties with night vision). Traffic accidents result in about 1.25 million deaths globally per year, with millions more sustaining serious injuries.[68] One study found that ride-sharing services have reduced accidents by an average of 6 percent.[69] Another found a 4 to 6 percent reduction specifically in drunk driving fatalities.[70] We estimate that reduced accident rates due to the expansion of digital mobility services could save $50 billion to $160 billion in economic terms—not to mention the incalculable value of reducing the human toll of accidents.

Beyond their effect on traditional taxi services, mobility services could have wider impact. Automakers are the biggest question mark as the calculus of car ownership changes, particularly for urban residents. While sales will likely continue to grow in absolute numbers, the shift toward mobility services could potentially halve the growth rate of global vehicle sales by 2030 (Exhibit 9).[71] In response, car manufacturers will likely need to diversify and lessen their reliance on traditional car sales. Many appear to be preparing for this future; partnerships are forming between traditional automakers and mobility service providers or other high-tech firms. Toyota, for example, recently invested an undisclosed amount

---

67 *Car data: Paving the way to value-creating mobility*, McKinsey Advanced Industries Practice, March 2016.

68 World Health Organization, 2013 data.

69 Angela K. Dills and Sean E. Mulholland, *Ride-sharing, fatal crashes, and crime*, Providence College working paper, May 2016.

70 Brad N. Greenwood and Sunil Wattal, *Show me the way to go home: An empirical investigation of ride sharing and alcohol related motor vehicle homicide*, Temple University, Fox School of Business research paper number 15-054, January 2015.

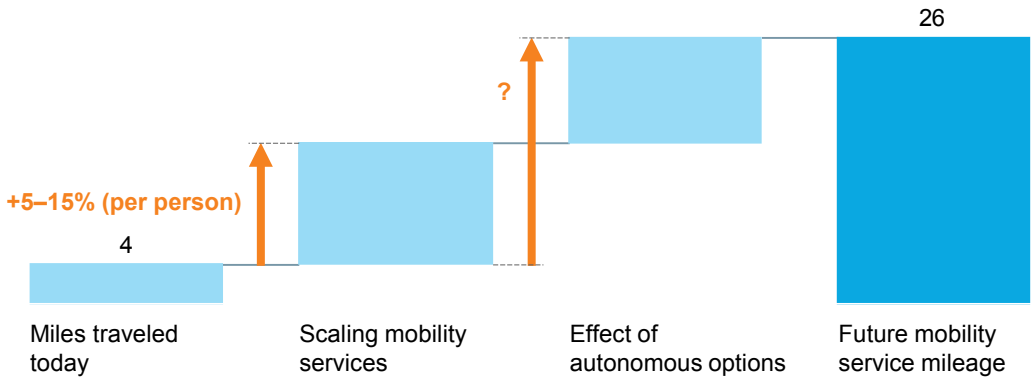
71 This draws on assumptions and similar findings from previous research by McKinsey & Company. See *Automotive revolution—perspective towards 2030*, McKinsey Advanced Industries Practice, January 2016.

in Uber.[72] Several other cross-sector alliances have formed in recent years, such as Volkswagen with Gett, GM with Lyft, and Fiat with Google.

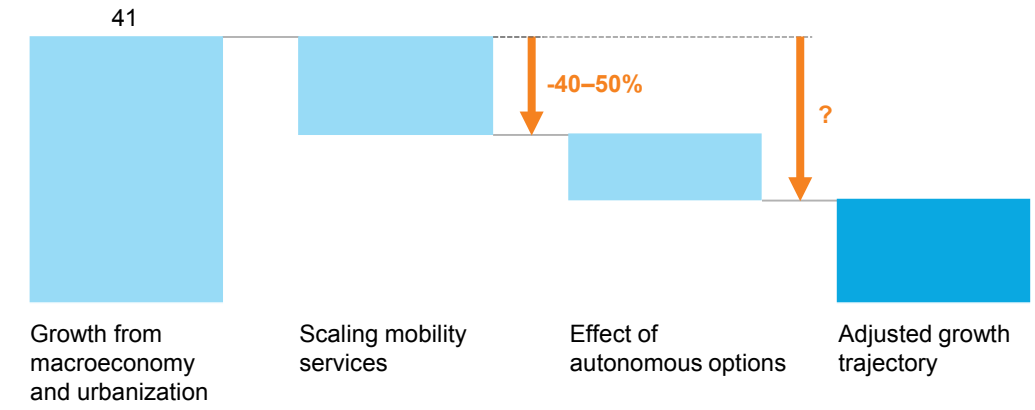
Exhibit 9

Increasing adoption of mobility services could lower the trajectory of global vehicles sales, potentially cutting the growth rate in half

Mobility service usage
Share of global mileage through 2030
%



New car sales
Expected increase in global new cars sold per year until 2030
Million



NOTE: Numbers may not sum due to rounding.

SOURCE: McKinsey Automotive 2030 report; Autos & shared mobility report, Morgan Stanley; McKinsey Global Institute analysis

Autonomous vehicles, which appear to be on the horizon, could accelerate this wave of change. When self-driving cars are added into the equation, supply and demand matching could improve even further since these vehicles can have higher utilization rates. Car pooling may increase, and the cost of urban transportation could plummet. On the flip side, the demand for car purchases could fall further, and many people who make a living as drivers (nearly two million in the United States alone with the majority being truck drivers) could be displaced.

The role of data and analytics in transportation is not limited to urban centers. It can also improve the efficiency of trucking routes and handoffs in the logistics industry. Rivigo has applied mapping technology and algorithms to improve logistics efficiency in parts of India.

[72] Douglas Macmillan, "Toyota and Uber reach investment, lease partnership," *The Wall Street Journal*, May 25, 2016.

Hyperscale matching platforms could be transformative in other areas as well

While urban transportation has been profoundly altered by mobility services, it is only one area where digital platforms have had a major effect. Retail, social networking, music, and even dating have been revolutionized by the introduction of hyperscale platforms. And because other markets have the same kind of inefficiencies that characterized the traditional taxi market before the advent of these services, they, too, could be transformed if platforms reach critical mass.

In energy markets, for instance, demand can fluctuate dramatically and frequently by time and by region. The current energy grid is ill-equipped to smooth out the spikes in peak demand with excess off-peak supply. But wider deployment of smart grid technology can address this inefficiency by using new sensor data to generate more dynamic matching of supply and demand, in part by allowing small, private energy producers (even individual homeowners) to sell excess capacity back to the grid. This technology is developing quickly: the United States alone has committed more than $9 billion in public and private funds toward smart grid technology since 2010.[73] In the Netherlands, some startups are using the peer-to-peer model to match individual households directly with small providers (such as farmers) who produce excess energy. Vandebron, for instance, charges a fixed subscription fee to connect consumers with renewable energy providers; in 2016, this service provided electricity to about 80,000 Dutch households.

The markets for certain types of short-term labor services are also being redefined. Driving passengers is only one of the many types of services now being offered through digital marketplaces. Others include household chores and errands, data entry, and simple coding projects. Conventional platforms (even digital job boards such as Craigslist) allow for static requests. But now platforms can match available workers with requests for their services on demand.

TaskRabbit, for example, serves 18 US cities, matching more than 70 percent of task requests with a local provider within five minutes. There are more than 30,000 "taskers" globally, and the average worker who participates takes on two or three jobs a day, five days a week.[74] Recent research from MGI has found that already some 15 percent of independent workers in the United States and Europe have used digital platforms to earn income.[75] The non-profit Samasource is seeking to bridge this market gap by breaking down larger digital projects into smaller discrete tasks that can be handled by remote workers in developing countries. As of 2016, almost 8,000 workers participated on this platform, increasing their earnings by more than three-and-a-half times.[76]

Platforms such as TaskRabbit and Samasource quickly match underutilized supply (people looking for work) with demand. This can have productivity benefits for businesses, while creating a new avenue for individuals who need work to generate income. Previous MGI research found that some 850 million people across seven major economies alone are unemployed, inactive, or working only part time. Previously they had few options for rejoining the workforce or increasing their hours, but these types of platforms increase the range of flexible options available to them.[77]

---

[73] *2014 smart grid system report*, US Department of Energy, August 2014.

[74] Company website.

[75] *Independent work: Choice, necessity, and the gig economy*, McKinsey Global Institute, October 2016.

[76] Company website.

[77] *A labor market that works: Connecting talent with opportunity in the digital age*, McKinsey Global Institute, June 2015; and *Independent work: Choice, necessity, and the gig economy*, McKinsey Global Institute, October 2016.

"RADICAL PERSONALIZATION" CAN TAILOR OFFERINGS TO THE INDIVIDUAL

One of the most powerful uses for data and analytics is micro-segmenting populations based on characteristics and preferences. By gathering and analyzing an increasing wealth of data, businesses get to know their customers at a deeper level. They can feed these insights back into their products and services and recommend additional purchases. Data and analytics take this capability to a new level, enabling what we refer to as radical personalization.

Exceedingly detailed data now enable finer levels of distinctions among individuals, a capability that paves the way to precise micro-targeting. Behavioral data gathered from diverse sources such as online activity, social media commentary, and wearables can yield personal preferences and insights. Broad data sets spanning large numbers of users or customers can continuously improve the experience. Amazon, for instance, uses algorithms to compare its interactions with one individual with results across large consumer data sets, generating targeted product recommendations from across its marketplace.

Radical personalization could be enormously valuable in many sectors, including education, travel and hospitality, media, retail, and advertising, as well as in the broader labor market (Exhibit 10). These are areas in which the good or service has a differentiated value for each individual, and preferences for distinct features or attributes affect the consumer's willingness to pay. The ability to customize on a mass scale creates the possibility of meeting an enormous spectrum of individual demands. Other digital technologies such as 3D printing can enable firms to execute on this strategy, since they allow for greater product personalization and do not require enormous economies of scale.

Exhibit 10

Radical personalization will be disruptive in areas where tailoring offerings to personal preferences and characteristics is highly valued

Data and analytics enablers

- Granular data enables finer levels of distinctions among individuals
- Outcomes and responses data allow businesses to estimate relationships between individual characteristics and improved value from customized goods/services

Industry preconditions

- The good or service has a differentiated value for each individual
- Mass customization creates possibility of meeting individual demands

Analytics will enable individually tailored products and services in these industries	
Health care	Tailoring interventions leads to precision wellness
Education	Individualized learning experiences based on existing skills, learning style, and interests
Labor market	Identifying an individual's skills and career goals to aid job matching and training
Travel and leisure	Customized travel experiences and recommendations
Media	Tailored and curated content
Retail	Shoppers directed to the right products for them at the right moment
Advertising	Ads targeted by time, location, and person to maximize potential sales

Matching complexity

SOURCE: McKinsey Global Institute analysis

Radical personalization has profound implications for the way health care is delivered

Standardized treatments do not work for every patient, given the complexity of each person's history and genetic makeup. But it is becoming possible to customize treatments and interventions (Exhibit 11). The data describing an individual patient's health are becoming much more granular with the declining costs of genetic sequencing, the advent of proteomics (the analysis of proteins), and the increasing number of sensors, monitors, and diagnostics that can provide a constant stream of real-time data. New approaches being developed (such as immunotherapy and CRISPR/Cas9 targeted genome editing) could potentially be tailored to maximize the chance of success given each person's makeup.

Today there is a growing push to integrate these new capabilities and vast patient data sets from electronic medical records into the delivery of care. Data and analytics may finally be ready to take off given the rapid growth in the volume of available data, the increasing recognition among all stakeholders that there is value in making better use of patient data, and changing incentives.

Advanced analytics could transform standardized disease treatments into personalized risk assessment, diagnosis, treatment, and monitoring. Some providers are already putting these capabilities to work and demonstrating the potential in clinical settings. Essentia Health, a health-care system in the US Upper Midwest, is using home monitoring for patients with congestive heart failure, reducing 30-day readmission rates to 2 percent, far below the national average of 25 percent.[78]

Radical personalization is partly about using granular, individual data to identify tailored treatments based on a patient's specific biomarkers, genetics, and behaviors. But it can also be put to broader use in transforming the system. In many countries around the world, and especially in the United States, a lack of information transparency and poorly aligned incentives create dysfunction. Providers are reimbursed for filling hospital beds and running procedures, and individuals lack the information to shop around and be informed consumers. Most patients enter the health system only when they have a disease. Care is focused in high-cost settings and not optimized for the patient experience or for value, in large part because the data that could be used to measure and monitor outcomes is not available to the parties that need it. Making better data available could help patients be more aware of their risk factors and take charge of their own health. Insurers can learn more about their customers and provide incentives for preventive measures. Hospitals and provider groups can be rewarded for outcomes rather than paid by the procedure, creating a system that steers patients toward the best place and time for intervention and connecting them with the right specialists. Using data to change incentives could have a huge impact in terms of both dollars saved and patient health.

[78] Elizabeth Stawicki, "Talking scales and telemedicine: ACO tools to keep patients out of the hospital," *Kaiser Health News*, August 2013.

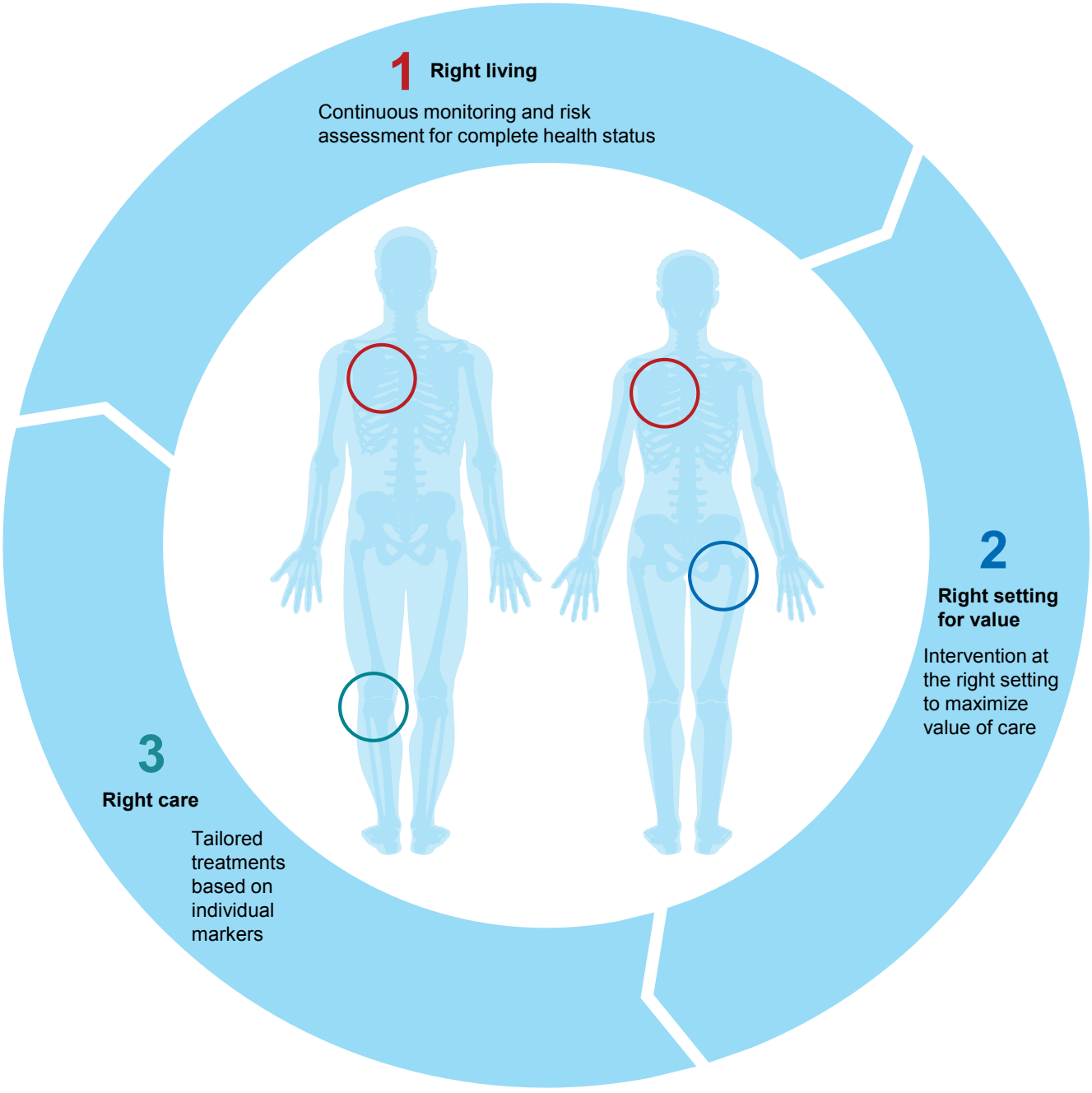
Exhibit 11

Radical personalization has the potential to transform how health care is delivered

Current state: Patients progress along a standardized disease treatment pathway

1. Patients only enter the health system when they have a disease
2. Patient case is focused in high-cost settings and not optimized for value
3. Physicians follow clinical guidelines for all patients with the same disease

Future state: Continuous monitoring and personalized treatment of patients at best place and time for intervention


1
Right living
Continuous monitoring and risk assessment for complete health status
2
Right setting for value
Intervention at the right setting to maximize value of care
3
Right care
Tailored treatments based on individual markers

Tangible results

- 5–9% lower national health expenditures
- Up to 1 year increase in health and life expectancy
- $200 per person increase in productivity

SOURCE: McKinsey Global Institute analysis

Radical personalization may reshape the health-care system

The advent of personalized care may alter the way stakeholders throughout the system operate, even as the dynamics will vary by country. The discussion below pertains largely to the US health-care system, but many of these changes will apply globally.

- **Providers:** To deliver truly personalized care, health-care networks and other providers will need to integrate data across EMR systems to get a complete view of a patient. It will take a vast data set of patient records to build smart clinical decision support tools. A great challenge for these systems and for the practice of medicine generally will be how to manage this constant flood of information and incorporate it into care. A doctor today sees a patient with asthma; by contrast, tomorrow's doctor will see an asthmatic patient who works outdoors, exercises daily, has certain genetic markers, and shows elevated expression of a few proteins. Physicians and regulators will need to consider carefully how to utilize such real-world evidence, which can enable them to put a greater focus on prevention and wellness. Realizing this potential may require a realignment of financial incentives away from fee-for-service reimbursement and toward a value-based model that emphasizes outcomes and prevention.

- **Payers:** Payers can use data and analytics to promote increased price transparency throughout the system. New partnerships among payers, providers, and pharmaceutical companies and pay-for-performance models may set the stage for this shift. Innovative partnerships, such as the one between Inova Health System and Aetna, have used data sharing in a value-based reimbursement model to provide a more integrated patient experience. Payers may become more involved in care management or encouraging their providers to do so. Adoption of these models has been slow, but the increasingly data-rich environment will enable better determination of which treatments are truly effective for certain patient profiles and at what cost. This can be beneficial not only in the US context but also in nationalized health-care systems.

- **Pharmaceutical and medical device companies:** On the R&D side, big data and advanced analytics can make predictive modeling of biological processes and drugs much more sophisticated, accelerating drug development. Scientific understanding of molecular biological processes will expand rapidly with a huge database of knowledge, and pharmaceutical companies can use genomic and proteomic data combined with millions of records on patient outcomes to design better therapies. Instead of aiming for the next blockbuster, pharma companies will be challenged to shift their business models to deliver tailored treatments for smaller, targeted patient populations. While oncology today is the clear focus for personalization, other treatment areas will follow as the right information becomes available.

Economic impact and disruption

Bringing data and analytics to health care has the potential to reduce costs and to allow people to enjoy longer, healthier, and more productive lives. We see the greatest potential in three aspects of health-care delivery: monitoring people to enable wellness and preventive care, directing them to the right care setting, and personalizing treatments. While much of the discussion surrounding personalized medicine has focused on the last dimension, monitoring technologies and analytics could have an even bigger impact if they can be used in combination with a realignment of incentive structures to design a system that emphasizes prevention and value-based care.

First, providers can use IoT technology and analytics to monitor patients remotely, timing interventions and making adjustments before a crisis occurs. This could be transformative in treating chronic conditions such as diabetes and cardiovascular and respiratory diseases and ensuring that patients are following recommended regimes. The use of these monitoring technologies dramatically reduces costs for the patient. They can also be used to change incentive structures. New business models can use these technologies, combined with other behavioral health interventions, to create a new focus on prevention, disease management, and wellness—addressing health before a person becomes a patient. For example, Discover Health, an insurer based in South Africa, tracks its consumers' food purchases and fitness behaviors, offering rewards and incentives for healthy behavior.[79]

Second, patients can be promptly directed to the right setting to receive diagnosis and care, shifting them away from higher-cost and sometimes higher-risk settings when possible. Analytics can enable proactive risk assessment to anticipate complications and avoid hospital stays, reducing overall health-care expenditures. Additional savings can come from providing patients, physicians, and insurers with more information on pricing and quality—a particular issue in the United States, where costs can vary sharply across providers for the same procedure. Evidation Health, for example, provides digital tools to stakeholders across the system that can aggregate data to determine the efficacy and cost-effectiveness of various interventions. Our analysis looked solely at the pricing transparency that will drive individual decisions about care, but the potential for broader population-based value assessments could be enormous. The National Institute for Health and Care Excellence (NICE) in the United Kingdom evaluates all medical interventions and treatments through a quality-adjusted life-year lens to determine cost-effectiveness, and many other systems are looking to make similar changes.

Finally, the last element of personalized medicine is identifying the right treatment for each patient. Clinical decision support systems, powered by artificial intelligence, will be able to comb through millions of patient records, genomic sequences, and other health and behavioral data to identify courses of treatment that are most effective for a particular individual with certain characteristics. Combining these insights with the costs of care could identify the most cost-effective treatment, while avoiding treatments that are ineffective for a given individual. This could maximize the efficacy of drugs, surgeries, and other interventions while reducing waste and harmful side effects. Researchers at University College London, for example, use supercomputer simulations to determine the best treatment for specific breast cancer mutations among 50 drug choices.[80] The integration of large patient data sets will enable the identification of new insights in patient care for very specific health and disease states.

---

[79] *Sustainable development report*, Discovery Health, 2015; Adrian Gore, "How Discovery keeps innovating," *McKinsey Quarterly*, May 2015.

[80] "SuperMUC enables major finding in personalized medicine," HPC Wire, June 20, 2016, available at www.hpcwire.com/off-the-wire/supermuc-enables-major-finding-personalized-medicine/.

1–2%
of GDP potential reduction in US health-care costs from greater adoption throughout the system

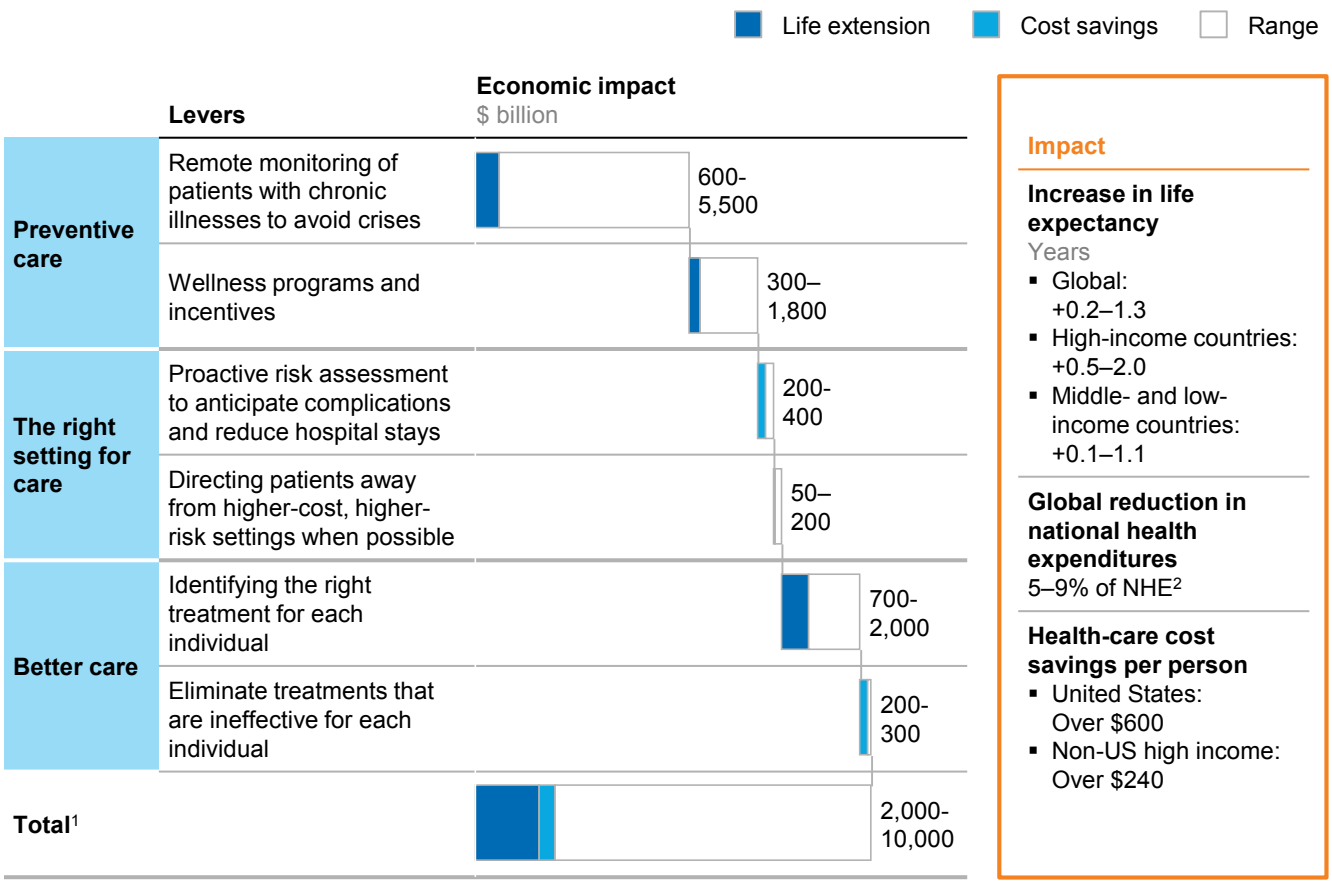
The potential benefits of radically personalizing the delivery of health care could be profound. The total impact could range from $2 trillion to $10 trillion on a global basis (Exhibit 12). The wide range is driven by a number of factors, including how rapidly health-care systems can adopt these techniques and the high degree of uncertainty around developing new treatments; the upper end of the range assumes an aggressive level of adoption within ten years.

The main driver of this impact is the potential effect on health and longevity. A more data-driven approach to health care could add a year of life expectancy in developed nations, which have the technology base in place to make this shift. Healthier lives also lead to increased productivity and quality of life, factors that are captured in our economic impact assessment.

The other major benefits are cost reductions. In the United States, where health-care spending is 18 percent of GDP, we estimate more than $600 of annual savings per person, which equates to 1 to 2 percent of GDP. Across other high-income countries, these savings would be closer to $200 per person, which translates to 0.5 to 1 percent of GDP.

Exhibit 12

Precision health care could drive better health outcomes, improve productivity and quality of life, and lower costs



1 Roughly 60% of the impact occurs in developed economies, with the remainder in emerging economies.
2 National health expenditures.
NOTE: Numbers may not sum due to rounding.

SOURCE: McKinsey Global Institute analysis

These estimates take into account only those changes in the actual delivery of health care driven by analytics-enabled radical personalization. It may therefore be understated, as we have not sized all the ways that payers, pharmaceutical companies, and others may adapt, or any additional ways that analytics may drive impact. Furthermore, we focus on existing technologies and do not account for other innovations on the horizon that could further bring down the cost of care or dramatically improve treatment of certain diseases.

Radical personalization could disrupt other domains

Radical personalization has relevance in a range of sectors. Education, for instance, has much in common with health care in that the individual characteristics of every participant in the system are unique and complex, but today relatively standardized services are provided to large groups of different individuals. Delivering personalized education will require data on individual students' learning style, the knowledge and skills they already possess, the goals they wish to pursue, and their progress. Educational environments can then be tailored to deliver the right mode of instruction through the most effective medium—and do so at the appropriate pace, helping each student master one topic before moving on to the next. Renaissance Learning, for example, has developed software-based assessments to help teachers adjust their approach for each student. Its tests are adaptive, meaning that the next question asked adjusts based on the student's performance on the previous one, as identified from a database of millions of past exams. A number of universities, including Arizona State University and Virginia Commonwealth University, are using analytics to improve graduation rates by identifying struggling students and intervening before they fall too far behind and drop out.[81]

Other industries where radical personalization could take hold include retail, labor markets, travel and hospitality, media, and advertising. In travel and leisure, customized experiences can be created, with recommendations for everything from restaurant menu items to tour guides. Google recently launched a travel app that will offer personalized recommendations, while WayBlazer, powered by IBM Watson, helps customers customize their trips based on their preferences. The media landscape is changing as Facebook's news feed algorithm determines which news articles will interest users based on a combination of personal and behavioral data.

INTEGRATING DATA FROM MULTIPLE SOURCES CAN HELP ORGANIZATIONS TAKE THEIR ANALYTICS CAPABILITIES TO THE NEXT LEVEL

The first step in creating value from data and analytics is ensuring access to all relevant data. This may be straightforward in theory, but it has proven to be much more difficult in practice. No matter what their industry, most large organizations have a department and business unit structure that tends to create silos. As a result, it is difficult to share information seamlessly across those internal boundaries. Different departments may have unique data systems that are not integrated or may even be incompatible with one another. This issue can be particularly acute in large companies with long histories and legacy systems (some of which endure from previous mergers and acquisitions), but it can occur even in relatively young companies. Managing data across these organizational barriers can be costly, and few companies have existing data systems that are conducive to running analytics on a large scale. Most organizations could also benefit from the ability to draw on new types of data from a range of external sources and combine it with their internal data.

Technologies such as data lakes are a promising concept for overcoming these issues. These types of new tools can simplify access across the enterprise by integrating all types of data into one easily accessible and flexible repository, and they are also useful for storing the vast real-time information flows coming from IoT sensors. Data lakes can integrate

[81] Danielle Douglas-Gabriel, "Colleges are using big data to identify when students are likely to flame out," *Washington Post*, June 14, 2015.

information from new sources relatively easily, even if the data are in different formats. They can incorporate structured data from relational databases, unstructured data from sources such as multimedia files and emails, or even information that appears in raw, inconsistent form. The entire organization can access the information contained in data lakes via an easy-to-use search layer. Because they support near-real-time data intake, data lakes can provide an ideal space for data scientists to perform discovery, build data-driven applications, and run major analytics projects. Companies such as Amazon, Facebook, and Goldman Sachs have adopted approaches such as these to achieve significant internal process improvements and create new products.

Massive data integration capabilities are valuable for industries with accessibility challenges or those in which data discovery has direct relevance for creating value. Two conditions in particular set the stage. First, industries are ripe if data silos are a longstanding issue—that is, if many companies have internal data residing in separate business units without adequate access by others. In these cases, multiple, redundant data sources blur the bigger picture, and different departments may reach varying conclusions about the same phenomena. Second, massive data integration capabilities are valuable in cases where combining unstructured and structured data from multiple sources (including new external sources) enables better decision making. A consumer product company, for example, can add comments about its brands posted on social media platforms to its existing sales reports and demographic analyses. Data lakes are one such tool for integrating new sources of data as they become available—and since they can store both structured and unstructured data, they can tap an enormous range of potential new sources.

Retail banking is ripe for massive data integration to change the nature of competition

Retail banking has always been data-rich, with stores of information on customers' transactions, financial status, and demographics. But few institutions have fully taken advantage of this data due to internal barriers that limit access across the organization, the variable quality of the data, and sometimes even an inability to see the value that data could generate.

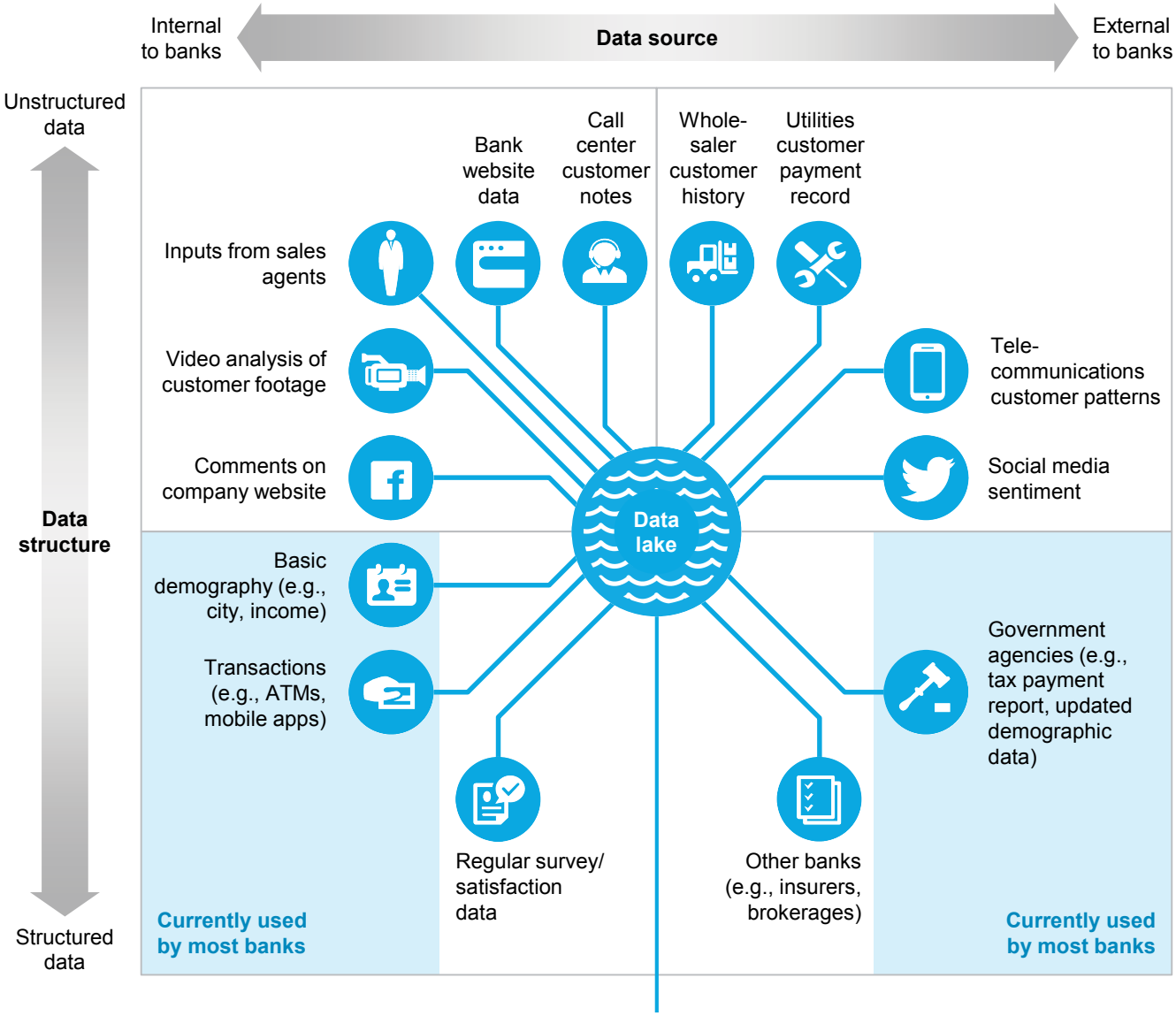
Surmounting these barriers is becoming critical now that a plethora of new data sources can be added to existing transaction records. These include social media posts, call center discussions, video footage from branches, and data acquired from external sources and partners. In addition, retail banks can partner with mobile phone operators and retailers to complement their view of each customer.

Adding analytics on top of all the data can enhance the customer experience, enabling banks to retain existing customers and attract new ones. Customers increasingly expect a personalized experience across all channels. They also want banking services to be available on their preferred platforms; a growing number are making payments directly via messaging apps, for instance. Integration tools such as data lakes help to provide a holistic view of each customer by combining different sociodemographic and behavioral data. One of the important applications for banks is the ability to improve risk and credit assessments by drawing on data from multiple sources for more accurate predictions.

Massive data integration repositories and the analytics they enable can also optimize banking operations and increase revenue (Exhibit 13). This is critical in an era when low interest rates are putting pressure on margins, regulatory and reporting requirements are growing more complex, and new digital native startups are introducing innovative business models into financial services.

Exhibit 13

Retail banks have opportunity to break their data silos, combining traditional and new data sources in data lakes

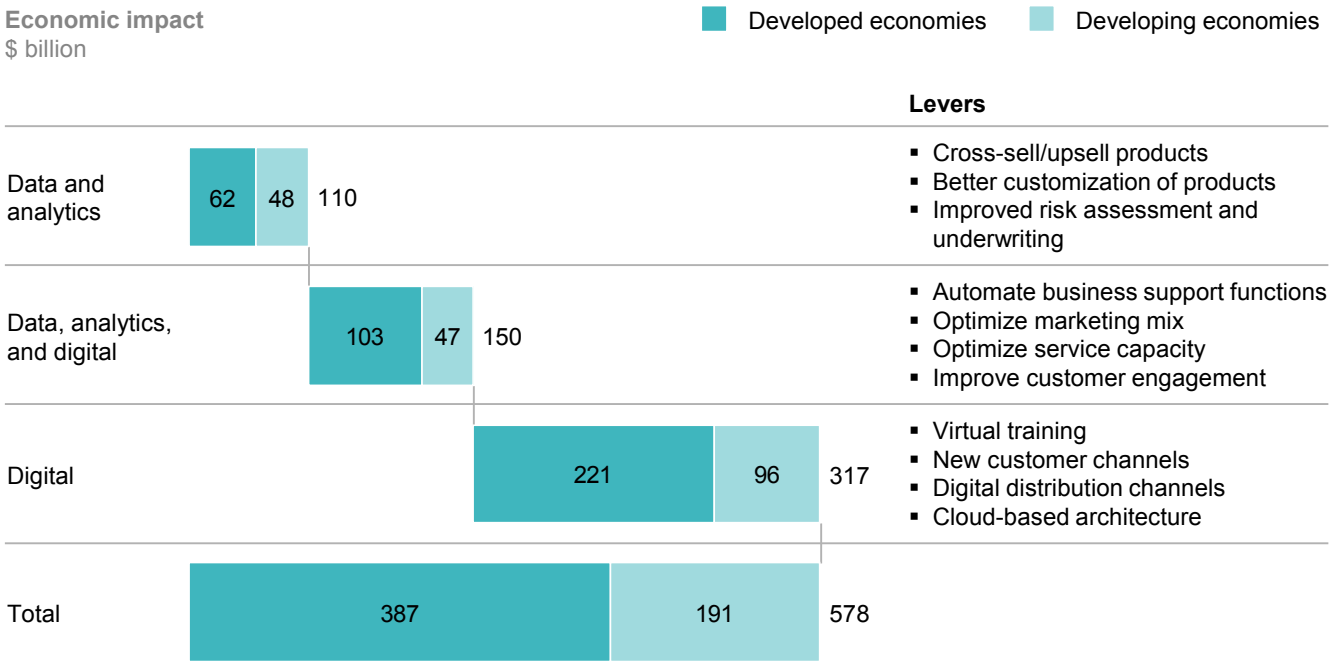

Internal to banks
Data source
External to banks
Unstructured data
Data structure
Structured data
Bank website data
Call center customer notes
Whole-saler customer history
Utilities customer payment record
Inputs from sales agents
Video analysis of customer footage
Comments on company website
Tele-communications customer patterns
Social media sentiment
Data lake
Basic demography (e.g., city, income)
Transactions (e.g., ATMs, mobile apps)
Government agencies (e.g., tax payment report, updated demographic data)
Regular survey/ satisfaction data
Other banks (e.g., insurers, brokerages)
Currently used by most banks
Currently used by most banks
▪ Stores practically unlimited amounts of data of any format and type
▪ Silos minimized, and single source of truth accessible by whole organization
▪ Offers an improved platform to run analytics and data discovery
▪ Transformation to the data lakes environment can be done gradually

SOURCE: McKinsey Global Institute analysis

Widespread adoption of analytics on top of massive data integration architecture has significant potential to create economic impact within the retail banking industry. Among the main mechanisms will be improved cross-selling and the development of personalized products. Analytics could have an especially large impact in reducing risk via better credit underwriting, credit monitoring, and improved collections. Retail banking is still undergoing digitizing many operation, and as this trend combines with a deeper use of data and analytics, new efficiencies can be realized in areas such as automating business support, targeting marketing more effectively, and improving customer engagement. The process of digitization itself will enable banks to generate more customer data while simultaneously taking advantage of lower-cost digital channels. Together these shifts could generate impact of some $400 billion to $600 billion annually, with roughly two-thirds of the impact realized in developed economies and the remainder in developing economies (Exhibit 14).

Exhibit 14

A data-driven transformation of the retail banking industry could generate roughly $400 billion to $600 billion in economic impact



NOTE: Revenue- and cost-saving levers are separated into three categories based on the size of the role analytics will play. See appendix for levers. Numbers may not sum due to rounding.

SOURCE: McKinsey Digital and Financial Services Practice; McKinsey Panorama global banking database; McKinsey Advanced Analytics Practice; McKinsey Global Institute analysis

As the retail banking industry becomes more data-driven, it is likely to take on an entirely new look. Talent capable of crunching the numbers, developing the algorithms, and combining those skills with an understanding of financial services and the regulatory landscape will be critical. As more banking services become digital, physical branches could be phased out, leading to substantial cost savings.

Three main types of ecosystem players could emerge. First are the solution and analytics innovators. This category includes many of the fintech players. These digital natives have deep capabilities, and they tend to focus on a particular market niche, with no intention of expanding into full retail banking offerings. This frees them from the constraints of being fully regulated banks. Betterment, for example, offers wealth management by robo-advisers, while ZestFinance combines vast amounts of data and machine learning to improve credit scoring.

Second are incumbent institutions that are the natural owners of significant amounts of data describing the financial position and behavior of their customers. This proprietary data gives them a significant advantage, but they could lose it if they lose the source of the data (that is, their customers) or if other players come up with ways to create equivalent or superior information by integrating from different sources.

Third, companies in other sectors can become part of the banking ecosystem if they bring in orthogonal data—such as non-financial data that provides a more comprehensive and granular view of the customer. These players may have large customer bases and advanced analytics capabilities created for their core businesses, and they can use these advantages to make rapid moves across sector boundaries, adding financial services to their business

lines. Alibaba's creation of Alipay and Apple's unveiling of Apple Pay are prime examples of this trend.

The move toward massive data integration and analytics may create room for new players and partnerships. Acquisitions and collaborations may be necessary for traditional retail banks to acquire the talent and capabilities they need to stay relevant. Santander UK, for example, has launched a partnership with Kabbage, a small business loan provider that uses machine learning to automate credit decisions. Data-sharing partnerships between banks and telecom companies have helped both sides develop new insights (especially around fraud, customer creditworthiness, and microsegments of customers) that otherwise would not have been possible.

Analytics-driven business models are emerging, such as Mint, an aggregator that utilizes data to create personal finance recommendations. Another new model involves peer-to-peer platforms. SoFi, for example, uses robust risk scoring through creative data sources and analytics to connect borrowers and investors directly; the company reports that it already has more than $6 billion in funded loans.[82]

Massive data integration has wide potential in other industries

In principle, there are no industries in which the ability to continuously integrate new sources of data of any format and quality would not generate improvements. In addition to retail banking, some industries stand out as having the characteristics that make them prime candidates for massive data integration capabilities to have a major impact.

In the public sector, for example, tools such as data lakes could break the silos that exist across various agencies and levels of government, although this will require rigorous data management and opening public data sets. It is often the case that systems and data are owned by different departments and functions, on a range of platforms and with differing taxonomies and access requirements. Fragmentation and the absence of a central owner for nationwide IT infrastructure and common components can make it hard to connect the internal "plumbing."[83] But massive data integration could create a more seamless experience for the end-user, whether that user is a government worker, a business, a citizen, or another intergovernmental office. Cities, for example, could link health and education data into planning their social welfare programs, while law enforcement could connect to external sources of data such as social media and weather conditions for better situational awareness.

Insurance is another industry in which incorporating more data from multiple sources and applying cutting-edge analytics can create value. The industry can rely on a wider array of information to assess risk more accurately. Companies can reduce fraud, improve pricing, and improve cross-selling by determining when people are going through a life change that may affect their product needs. Data mining also helps in creation of new products.

Manufacturing and processing industries can also benefit from storing all of their reports and logs in a single repository where patterns can be extracted when enough historical data are collected. This can be an especially valuable approach for collecting and analyzing the vast information flows generated by IoT sensors.

[82] "SoFi surpasses $6 billion in funded loans, bolsters leadership team," company press release, December 17, 2015.

[83] Cem Dilmegani, Bengi Korkmaz, and Martin Lundqvist, "Public-sector digitization: The trillion-dollar challenge," McKinsey.com, December 2014.

DATA AND ANALYTICS CAN TRANSFORM DISCOVERY AND INNOVATION

Innovation is one of the components of productivity growth. It can be applied to both processes and products. Process innovation enables an organization to put its resources to the most efficient use, while product innovation is about developing new products or a better customer experience. Throughout history, innovative ideas have sprung from human ingenuity and creativity. But what if data and algorithms could support, enhance, or even replace human ingenuity in some instances? Data and analytics techniques are increasingly being used to discover patterns in huge amounts of diverse data and to generate hypotheses.

In the realm of process innovation, data and analytics are helping organizations determine how to structure teams, resources, and workflows. High-performing teams can be many times more productive than low-performing teams. Understanding this variance and how to build more effective collaboration is a huge opportunity for organizations. This involves looking at issues such as complementarity of skills, the optimal team size, whether teams need to work together in person, what past experience or training is important, and even how their personalities may mesh. Data and analytics can be used to generate new hypotheses through finding new patterns that may not have even occurred to managers. Vast amounts of email, calendar, locational, and other data are available to understand how people work together and communicate, and all of these data can lead to new insights about improving performance.

In product innovation, data and analytics can transform research and development in areas such as materials science, synthetic biology, and life sciences. Leading pharmaceutical companies are using data and analytics to aid with drug discovery. Data from a variety of sources could help to suggest the chemical compounds that could serve as effective drug treatments for a variety of diseases. Furthermore, with huge amounts of data to sort through and nearly infinite possible combinations of features, deep learning techniques help to narrow the universe of possible combinations in a "smart" way, leading to discoveries. One team of scientists at Carnegie Mellon University used data and machine learning to predict the results of experiments without actually having to perform them, thus reducing the number of tests by 70 percent.[84] In another example, AstraZeneca and Human Longevity are partnering to build a database of one million genomic and health records along with 500,000 DNA samples from clinical trials. The associations and patterns that can be gleaned from those data could prove to be immensely valuable in advancing scientific and drug development breakthroughs.[85]

Companies are also using data and analytics to improve the online user experience. They can experiment with variables such as the optimal design and placement of content on a web page, smart recommendations, user journeys, and various types of A/B testing. Deep learning is also getting better at creating art, logos, and other design content that could affect creative fields.

Applying data and analytics to product, process, and design innovation is very much in the early stages—but some organizations are pushing the boundaries in these applications. If they are able to develop these capabilities to differentiate themselves, they could move further ahead of the competition.

---

[84] Jake Hissitt, "Machine learning for discovery: How machine learning algorithms are revolutionizing pharma," Innovation Enterprise, April 22, 2016; and Chris Wood, "Machine-learning robot could streamline drug development," New Atlas, February 10, 2016.

[85] "Human Longevity Inc. announces 10-year deal with AstraZeneca to sequence and analyze patient samples from AstraZeneca clinical trials," company press release, April 2016.

ANALYTICS CAN OVERCOME HUMAN LIMITATIONS TO IMPROVE THE SPEED, ACCURACY, CONSISTENCY, AND TRANSPARENCY OF DECISIONS

Decision making is fundamental to everything, but when faced with important questions, most people ultimately rely on gut instinct. Today, however, data and analytics can turn decision making into a more data-driven exercise, helping to overcome human limitations and biases (Exhibit 15).

Faulty assumptions and biases lead many hiring managers to screen out "non-traditional" job seekers who might have tremendous natural ability. In other types of decisions, people tend to give weight to data points that back up their original hypothesis rather than fully considering those that contradict their working theory (a phenomenon known as confirmation bias). Or they may prioritize data points that support continuing the status quo or avoiding risk (stability bias). The use of data can obviate some of these biases, making the decision-making process more transparent and evidence-based.

Exhibit 15

Data and analytics can overcome human limitations in decision making

Preconditions
- Human biases and heuristics are predominant in decision making
- Human error and physical limitations lead to mistakes and lost value

Smart cities	A more effective city planning and zoning process that better captures the needs of the community
Insurance and risk	Increased behavioral data from sensors and other sources leads to better policy pricing and risk assessment
Health care	Algorithms can prevent human errors in health care, avoiding misdiagnoses, drug interactions, and incorrect dosages
Criminal justice	Smart algorithms can improve sentencing and parole decisions, while predictive policing can better direct law enforcement to crime
Labor markets	Job seekers and employers will have greater transparency that enables them to make better matches

SOURCE: McKinsey Global Institute analysis

Other human limitations in decision making can also be overcome through the use of data and analytics. Today there is a flood of data from new sources, such as IoT sensors, digital games, and social media images. While humans can reach information overload, automated algorithms can weigh a vast amount of data. Automated algorithms and decision-support tools can help to avoid errors or lapses that have serious consequences. Self-driving cars, for example, can continuously monitor their surroundings without falling prey to fatigue or distraction, which can cause human drivers to have accidents.

Analytics can improve four aspects of decision making:

- **Speed/adaptability.** Machines and algorithms can react in an instant. This a critical capability when real-time responses are needed. Cities, for example, need to move quickly to respond to crime, natural disaster, or the failure of infrastructure systems. In transportation, safety is a key concern, and autonomous vehicles could respond faster than human drivers to sudden obstacles such as a deer crossing the highway. Investors similarly need to be able to respond quickly to swings in financial markets, and many institutions have already adopted algorithmic support to automate buying and selling. It should be noted, however, that high-frequency trading has sparked controversy, with critics asserting that it harms the stability of financial markets.

- **Accuracy.** Predictive models, when provided with the right data, can give a clearer view into the future, leading to more effective and judicious use of resources. Accuracy is of special importance when the consequences of small deviations are high. In health care, choosing the right medication and dosage is critical for treating a life-threatening disease.

- **Consistency/reliability.** Machines and algorithms are generally predictable and reliable. They do not tire, miss data points, or look at the same piece of information and draw varying conclusions each time. This consistency can reduce costly errors. Building information modeling systems, which create digital models of structures under design, employ analytics to help engineers and architects make the many decisions that go into a large-scale project, ensuring structural integrity and minimizing rework. Advances in robotics can reduce errors in manufacturing, shipping, labeling, and many other processes in the supply chain.

- **Transparency.** When two parties in a transaction have different sets of information, it can lead to suboptimal decision making. Transparency allows for decisions to be reviewed and improved upon in the future. Patients do not always have accurate information on what medical procedures cost at different facilities; this impedes their ability to make cost-effective choices, ultimately driving up costs for themselves, insurers, and employers. Services such as Castlight are working to provide transparency and enable employers that offer coverage to craft incentives for employees to help in holding down costs. Transparency in the public sector has immense value to citizens, empowering them to hold government accountable. More transparency could improve decisions regarding the allocation of scarce resources in education or health care, where to build parks or casinos, or who is issued permits. In the legal system, decisions could become more impartial when they are informed by as much data as possible.

Below we look at example domains in which these dimensions of decision making are crucial and consider how analytics capabilities could be transformative. In many cases, this could create societal benefits in addition to economic value. However, it is also important to consider the fact that enhanced decision-making capabilities could lead to job displacement. Algorithms may allow machines to replace humans in many contexts—and as autonomous robots gain the ability to move and react in the physical world, the number of settings where human labor could conceivably be replaced will expand. But algorithms will not replace people everywhere they are used; in many cases, they are best used to support and complement human judgment rather than to substitute for it.

Faster decision making: Smart cities

10-20%
potential reduction in traffic congestion

Automating complex decision making could enable smart cities to become more efficient and better able to respond to changing environments in real time. Transportation and utilities in particular are two areas of urban management in which rapid decision making is crucial.

Smart transportation systems utilizing the IoT and automated algorithms can enable more seamless traffic flows, reduce waits on public transportation systems, and make parking availability fully transparent to prevent circling and congestion. Some cities have begun to deploy technologies that can produce these benefits. In Singapore, sensor data is used to predict traffic congestion in real time and adjust tolls to limit jams. In Copenhagen, road sensors detect approaching cyclists and turn signals green.

Utilities can use sensors to be more responsive to changes in demand or system stresses. Energy, waste, and water needs can change minute by minute, and fully automated algorithms can allow for efficient pricing and distribution. Tools such as smart grids enable utilities to deliver renewable energy alongside conventional sources.[86] Kenya Power is migrating tens of thousands of businesses to smart meters to combat tampering and cut the cost of reading meters while minimizing outages.[87]

We estimated the size of the economic potential associated with a selection of analytics applications in smart cities around the world.[88] In transportation management, we find that a number of uses could produce $200 billion to $500 billion in economic impact through reducing traffic congestion by about 10 to 20 percent. These include centralized and adaptive traffic control, smart parking meters and pricing, schedule management of buses and trains, congestion pricing in express lanes on roads, and monitoring public transit for maintenance needs.

Using analytics to make utilities more responsive in real time could create global impact of some $20 billion to $50 billion. This could come about through a combination of structural monitoring (for example, installing sensors on streetlights and bridges), smart meters for energy and water demand management, and automating power distribution and substations. These initiatives could decrease water consumption by up to 5 percent, produce cost savings of up to 20 percent in maintenance and labor, and reduce outages by more than 35 percent.

Accuracy: Predictive analytics in insurance

Predictive analytics can make forecasts and projections more accurate and precise. It can help businesses with a wide array of problems: predicting which customers will prove to be most valuable, preventing employee attrition, scheduling maintenance in time to prevent breakdowns, and detecting fraud. Many types of orthogonal data can be brought to bear in these contexts.

These capabilities have valuable applications for insurers. Over time, insurance companies have sought ways to collect additional data that they believe could have predictive power—and today they finally have the analytics ability to put that information to work in revolutionary ways. For example, behavioral data from sensors embedded in cars can provide insurers much more direct information about an individual's driving patterns and risk than his or her educational attainment, age, or make of car. Some insurers incorporate credit scores because of empirical evidence that people who pay their bills on time are also better drivers.

Companies offering property insurance are using sensors on water pipes or in the kitchen to identify predictors of pipe leaks or breaks, flooding, or fires, and to warn the homeowners. Similarly, behavioral data can transform life insurance models. Insurers can use these data to better price and customize coverage options—or even to warn customers to take preventive actions. The benefits of data and analytics should accrue to both insurers (in the form of reduced claims) and individuals (in the form of loss prevention). One UK insurance company that used vehicle sensors reported that better driving habits resulted in a 30 percent reduction in the number of claims, while another reported a 53 percent drop in risky driving behavior. Extrapolating that type of result across all auto insurers yields potential impact on the order of $40 billion. Summing up the benefits across other insurance industries and lives saved would indicate economic benefits in the hundreds of billions of dollars globally.

---

[86] Adrian Booth, Niko Mohr, and Peter Peters, "The digital utility: New opportunities and challenges," McKinsey.com, May 2016.

[87] Joshua Masinde, "Kenya Power counts on smart meters to cut costs and deal with tampering," *Daily Nation*, July 23, 2015.

[88] *The internet of things: Mapping the value beyond the hype*, McKinsey Global Institute, June 2015.

Consistency: Reducing errors in drug prescriptions

In medicine, drug interactions and incorrect dosages often have life-or-death consequences—and treatment decisions are often made under time-constrained and pressure-packed circumstances. Advanced analytical support tools and automated decision making have the ability to recommend the right drug at the right dosage and flag potential allergies or multi-drug reactions. In the United States, a recent report found that preventable medical errors were the third leading cause of death.[89] Another study found that nearly 5 percent of hospitalized patients experience adverse drug events, which include incorrect drug choices and dosages.[90]

The impact of reducing medical errors could be huge. It would encompass not only the direct additional costs to the medical system to correct the errors but also the lost productivity of patients—and most important, it could save lives. One landmark study in 2000 estimated that anywhere from 44,000 to 98,000 Americans die each year as a result of preventable medical errors.[91] Later research building on that study and applying a quality-adjusted life-year analysis estimated that the associated costs could be as high as $100 billion annually in the United States alone.[92] Assuming that analytics could reduce such errors by approximately half, the total impact in high-income countries could approach $200 billion to $300 billion.

Transparency: Transforming hiring and making labor markets more efficient

The global labor market can be thought of as a $30 trillion market that is rife with inefficiencies. At the top of the list of those inefficiencies are information failures at both the micro and macro levels. Employers do not often have a very clear sense of where to find the talent they need or what skills any given candidate might have. Meanwhile, job candidates have poor information on which jobs and skills are in demand and the pathway to get there. They also do not have clarity on which educational and career choices would be best for them. Decisions on all sides of the market can be ad hoc and based on faulty assumptions. Hiring sometimes comes down to having attended the right school or sharing the same interests as an interviewer. While these data points can provide insights, they are not always good proxies for the key factors firms are concerned about, namely worker productivity and likelihood of retention. Data and analytics can enable more informed decision making on all sides.

New analytics tools that incorporate online tests and games can give employers data markers indicating whether candidates have the qualities and skills that predict high performance and retention.[93] Many new players are developing solutions that can support smarter and more data-driven hiring processes. HireIQ, for example, provides software to digitize the interview process and apply predictive analytics to the results. Joberate enables companies to see real-time employment trends and competitive intelligence, while Knack uses gamification to test for skills and help match workers and jobs. But human resources departments will need an infusion of analytical talent to be better equipped to manage in a more data-driven world.

---

89 Martin A. Makary and Michael Daniel, "Medical error: The third leading cause of death in the US," *BMJ*, May 2016.

90 "Medication errors," Patient safety primer, Agency for Healthcare Research and Quality, US Department of Health and Human Services, available at https://psnet.ahrq.gov/primers/primer/23/medication-errors.

91 Linda T. Kohn, Janet M. Corrigan, and Molla S. Donaldson, eds., *To err is human: Building a safer health system*, Committee on Quality of Health Care in America, Institute of Medicine, National Academy Press, 2000.

92 C. Andel, S. L. Davidow, M. Hollander, and D. A. Moreno, "The economics of health care quality and medical errors," *Journal of Health Care Finance*, volume 39, number 1, fall 2012.

93 Susan Lund, James Manyika, and Kelsey Robinson, "Managing talent in a digital age," *McKinsey Quarterly*, March 2016.

Indeed, LinkedIn, Monster, and other online talent platforms are giving job seekers and employers alike better transparency.[94] In the future, employers and job seekers could gain greater visibility into real-time supply and demand, the wages associated with various jobs, specific skills, goals, and abilities, and the value (and costs) of different educational programs.

Beyond altering hiring and job hunting, this new transparency could help educational institutions better prepare students for the job market. Education providers can become more responsive to the demand for skills in the labor market. Universities can adjust their curricula to respond to those changes, while secondary schools could incorporate job preparation and discernment driven by data. MGI's previous research estimated that across seven major economies alone, some $89 billion of annual spending on tertiary education is rendered ineffective because students pursue paths that leave them unemployed or underemployed.[95] Meanwhile, earlier assessments could allow for more personalized and predictive educational planning that incorporates a return on investment analysis.

We sized the impact that these changes could have on the improved productivity of workers and the operational cost savings for an organization.[96] Higher productivity can come from better matching, training, and maximizing of talent, which could lead to $500 billion to $1 trillion in value worldwide. Meanwhile, firms can save money by reducing attrition and as well as creating time savings on recruiting. Such effects could total $100 billion to $150 billion in cost savings worldwide.

•••

Data and analytics are already shaking up multiple industries, and the effects will only become more pronounced in the years ahead as more players adopt and combine these models to affect even more industries. But the capabilities described in this chapter are only the tip of the iceberg. An even bigger wave of change is looming on the horizon as deep learning reaches maturity, giving machines unprecedented capabilities to think, problem-solve, and understand language. The following chapter looks at these groundbreaking developments and considers the enormous opportunities and risks they could pose.

---

[94] *A labor market that works: Connecting talent with opportunity in the digital age*, McKinsey Global Institute, June 2015.

[95] Ibid.

[96] See *A labor market that works* for more details on our assumptions regarding the impact of data and analytics in a number of human resources functions for a representative company in a few industries.

© Paper Boat Creative/Getty Images

5. DEEP LEARNING: THE COMING WAVE

Can a machine think? While the notion of artificial intelligence has been around for decades, recent advances in algorithms and processing power, combined with the exponential growth in available data, are enabling the creation of machines with unprecedented capabilities. While these technologies might not redefine what it means to "think," they are starting to perform activities long thought to be the sole purview of humans—sometimes at higher levels of performance than people can achieve.

Conventional software programs are hard-coded by developers with specific instructions on the tasks they need to execute. While this works well in many situations, this type of top-down instruction manual has limitations. The human programmer cannot code provisions that account for every possible state of the world. If the environment changes, the software programs will malfunction or cease to be relevant. By contrast, it is possible to create algorithms that "learn" from data and, if necessary, adapt to new circumstances without being explicitly reprogrammed. The concept underpinning machine learning is to give the algorithm "experiences" (training data) and a generalized strategy for learning, then let the algorithm identify patterns, associations, and insights from the data—in short, to train the system rather than program it.

What was once science fiction is now rapidly advancing science—and will soon be a core business capability. Systems enabled by machine learning can provide customer service, manage logistics, analyze medical records, or even write a news story. Deep learning, a frontier area of research within machine learning, uses neural networks with many layers (hence the label "deep") to push the boundaries of machine capabilities. Data scientists working in this field have recently made breakthroughs that enable machines to recognize objects and faces, to beat humans in challenging games such as chess and Go, to read lips, and even to generate natural language. Digital giants such as Google, Facebook, Intel, and Baidu as well as industrial companies such as GE are leading the way in these innovations, seeing machine learning as fundamental to their core business and strategy.

The potential uses of machine learning are remarkably broad. The value potential is everywhere, even extending into sectors that have been slow to apply data and analytics. As applications of this technology are adopted, they could generate tremendous productivity gains and an improved quality of life. But they could also unleash job losses and other disruptions, not to mention thorny ethical and societal questions that will have to be addressed as machines gain greater intellectual capabilities.

To give an idea of how this transformation could play out, we begin by exploring the potential impact of machine learning through two lenses. First, we investigate which business problems across 12 industries could be solved by machine learning. Second, we examine which work activities currently performed by people could potentially be automated through machine learning and how that could play out across occupations.

We caution that this is an initial and broad exploration rather than a deep investigation into specific industries and use cases. Furthermore, this is not meant to be a comprehensive list of every possible application of machine learning. It could be harnessed to tackle problems beyond the boundaries of the industries we analyzed, such as climate change. It could also find a role in daily social interactions that have nothing to do with commerce. The findings here are meant to set the stage for many avenues of interesting future research. All that

being said, one of the most striking of our initial findings is the incredibly wide applicability of machine learning techniques to many industry problems and individual occupations.

MACHINE LEARNING HAS APPLICATIONS ACROSS MANY INDUSTRIES, AND THERE ARE SUBSTANTIAL BUSINESS OPPORTUNITIES

The most immediate question for businesses is how machine learning algorithms could be applied and where the biggest impact is likely to occur. In some domains this technology is already creating value, while in others the potential is still uncertain. To identify high-impact arenas, we categorized what types of problems machine learning techniques are best suited to solve. We then identified specific use cases and conducted a survey of industry experts to gauge which ones they view as most valuable. Mapping these opportunities against the availability of data yields a set of use cases that combine high potential impact with the richest data.

What kind of problems is machine learning best suited to tackle?

Machine learning encompasses a number of algorithms or techniques that recognize patterns and associations in huge amounts of complex data. Techniques such as regression, support vector machines, and k-means clustering have been in use for decades. Others, while developed previously, have become viable only now that vast quantities of data and unprecedented processing power are available. Falling into the latter category are artificial "neural networks," which are inspired by the connectivity of neurons in the human brain. Early versions, in a world with far less processing power and available data, were built with only a few layers. But now an abundance of processing power (consistent with Moore's Law) and data has enabled researchers to train "deep" neural networks with ten or more layers—changing the game for uses such as image recognition. Data scientists have recently made breakthroughs using deep learning to recognize objects and faces and understand and generate language.

Reinforcement learning is another machine learning technique used to identify the best actions to take now in order to reach some future goal. These type of problems are common in games and can be useful for solving dynamic optimization and control theory problems—exactly the type of issues that come up in modeling any complex system in fields such as engineering and economics. Reinforcement learning algorithms that use deep neural networks ("deep reinforcement learning") have made breakthroughs in mastering games such as chess and Go.

All machine learning algorithms require large amounts of training data ("experiences") in order to learn. They recognize patterns in the training data to develop a "model" of the world being described by the data. Reinforcement learning is slightly different than other techniques in that the training data is not given to the algorithm, but rather, is generated in real time via interactions with and feedback from the environment. But in all cases, as new training data comes in, the algorithm is able to improve and refine the model. This process is especially suited for solving three broad categories of problems: classification, prediction/estimation, and generation (Exhibit 16).

First, tackling classification problems involves making observations about the world, such as identifying objects in images and video or recognizing text and audio. Classification also involves finding associations in data or segmenting data into clusters due to associations. (Customer segmentation is a classic example of this.) Second, machine learning can also be used for predictions such as estimating the likelihood of events and forecasting outcomes. Lastly, machine learning can be used to generate content, from interpolating missing data to generating the next frame in a video sequence.

Exhibit 16

Machine learning can help solve classification, prediction, and generation problems

Classification	Classify/label visual objects	Identify objects, faces in images and video
	Classify/label writing and text	Identify letters, symbols, words in writing sample
	Classify/label audio	Classify and label songs from audio samples
	Cluster, group other data	Segment objects (e.g., customers, product features) into categories, clusters
	Discover associations	Identify that people who watch certain TV shows also read certain books
Prediction	Predict probability of outcomes	Predict the probability that a customer will choose another provider
	Forecast	Trained on historical data, forecast demand for a product
	Value function estimation	Trained on thousands of games played, predict/estimate rewards from actions from future states for dynamic games
Generation	Generate visual objects	Trained on a set of artist's paintings, generate a new painting in the same style
	Generate writing and text	Trained on a historical text, fill in missing parts of a single page
	Generate audio	Generate a new potential recording in the same style/genre
	Generate other data	Trained on certain countries' weather data, fill in missing data points for countries with low data quality

SOURCE: McKinsey Global Institute analysis

What are the best business opportunities for machine learning?

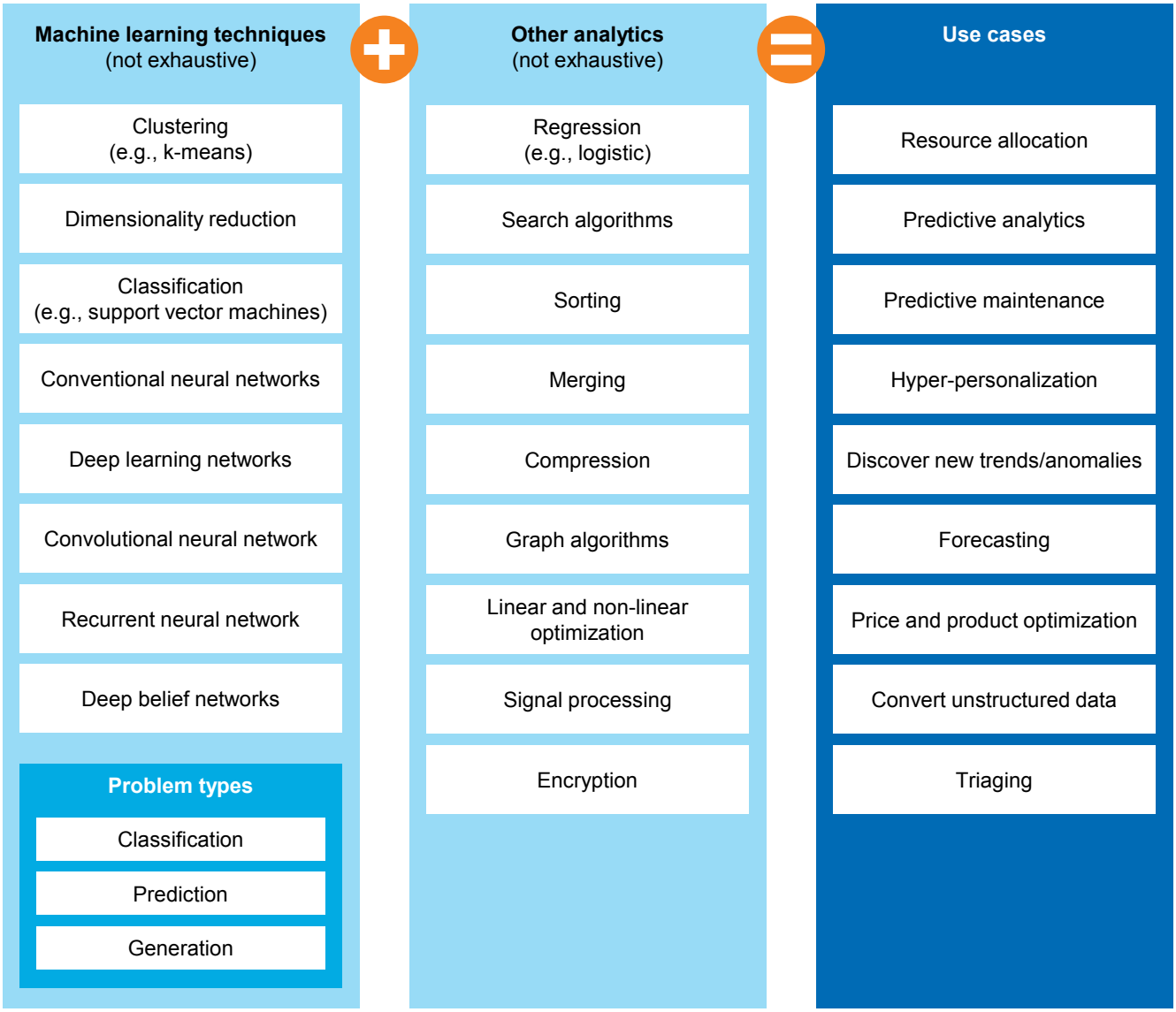
Combined with conventional optimization and statistical methods, machine learning can be useful in a number of settings (Exhibit 17).

After interviewing nearly 50 industry experts, we identified more than 300 specific use cases across 12 industries. We culled that list to the top ten use cases in each industry based on the size of the opportunity. We then surveyed a wider group of more than 600 experts across a variety of industries to determine where they saw the greatest potential to create value. Results from this survey suggest that the opportunity for business impact is broad. When we asked experts to rank individual use cases in their industry, each of the 120 use cases was named as being one of the top three most valuable in its industry by at least one industry expert.

However, looking at the value creation opportunity is only part of the picture. As discussed above, machine learning algorithms require a large amounts of data in order to be trained and effective. For example, improving hiring matches would have tremendous value for creating a more efficient labor market—and machine learning techniques are well suited to making more accurate matches. But the quantity and richness of data collected on candidates is quite limited; the typical individual has far fewer interactions in the labor market than they do on social media or in the course of online shopping. This potential of machine learning in the labor market could be constrained by this factor.

Exhibit 17

Machine learning can be combined with other types of analytics to solve a large swath of business problems



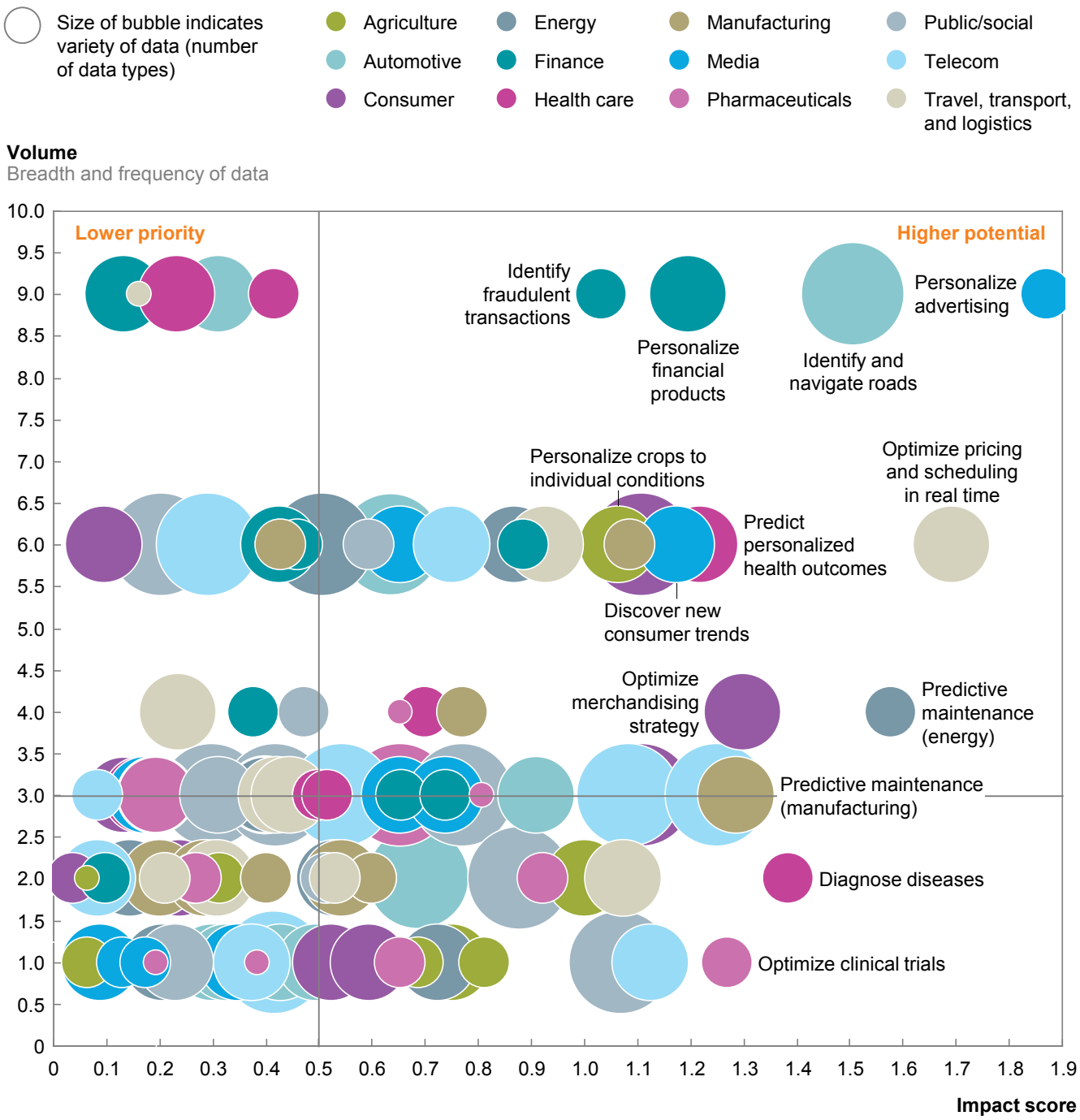
SOURCE: McKinsey Global Institute analysis

Our analysis filters business use cases by impact potential and by data richness. We plotted the top 120 use cases across 12 industries in Exhibit 18. The y-axis shows the volume of available data (encompassing its breadth and frequency), while the x-axis shows the potential impact. The size of each bubble reflects the diversity of the available data sources.

The industry-specific uses that combine data volume and frequency with a larger opportunity are in the top right quadrant of the chart. These represent areas where organizations should prioritize the use of machine learning and prepare for a transformation to take place. Some of the high-opportunity use cases include personalized, targeted advertising; autonomous vehicles; optimizing pricing, routing, and scheduling based on real-time data in travel and logistics; predicting personalized health outcomes; and optimizing merchandising strategy in retail. (See the technical appendix for a detailed scoring of use cases by industry, as well as a discussion of methodology.)

Exhibit 18

Machine learning has broad potential across industries and use cases



SOURCE: McKinsey Global Institute analysis

These use cases in the top right quadrant fall into four main categories. First is the radical personalization of products and services for customers. Second is predictive analytics, which involves not only forecasting but also uses such as anticipating fraud and bottlenecks or diagnosing diseases. The third high-impact category for machine learning is strategic optimization, and the fourth is real-time optimization in operations and logistics. Below we will examine each of these in turn.

The radical personalization of goods or services is one of the most exciting capabilities of machine learning, and it is one that many industries have yet to exploit. Chapter 4 discussed the enormous potential associated with bringing this capability to health care, but it also has tremendous scope in sectors such as consumer packaged goods, media, finance, and education—all areas where organizations can create huge value by tailoring their offerings to suit the preferences, characteristics, and needs of each person they serve. These industries collect a wealth of data about individuals, not only because they have large numbers of users but because they have repeated interactions with each one over time. Machine learning is well suited to identify patterns in large and granular data sets to segment populations all the way down to the individual level. Owners of this data will occupy prime positions in markets where customers are willing to pay for personalization.

Consider the potential impact of radical personalization in several specific areas. Some media organizations, for instance, are beginning to use this technology to deliver personalized content and advertising. Netflix's recommendation engine currently influences about 80 percent of content hours streamed on its platform. One study issued by the company estimates that using personalization has increased subscriber retention and engagement to such a degree that it is worth some $1 billion annually to the company.[97] (See Chapter 4 for more on this topic.)

Predictive analytics is another type of use with tremendous potential across almost all industries (see box at left). In these applications, machine learning helps classify customers or observations into groups for predicting value, behavior, risk, or other metrics. It can be used to triage customer service calls; to segment customers based on risk, churn, and purchasing patterns; to identify fraud and anomalies in banking and cybersecurity; and to diagnose diseases from scans, biopsies, and other data.

Industry applications of predictive analytics

Health care: Diagnose diseases from scans, biopsies, audio, and other data

Telecom: Predict the lifetime value and risk of churn associated with individual customers

Pharmaceuticals: Predict outcomes from experiments to reduce R&D costs and time to market

Finance: Evaluate customer credit risk more accurately, with faster and less biased underwriting decisions

Public sector: Optimize public policy decisions (for example, housing policies) to take into account a greater set of complex interactions and potential consequences

One media company, for example, used machine learning techniques to discover the factors that were most predictive of customer churn and identified the 2 percent of customers causing almost 20 percent of overall churn. Credit card issuers can identify the most important behaviors and characteristics that predict whether a consumer will apply for a credit card, insights that directly translate into targeting likely potential customers. A large retail bank in the United Kingdom used machine learning algorithms to identify fraudulent transactions with over 90 percent accuracy. At one large auto insurer, high accident rates for new client policies suggested that claims were being filed for pre-existing damage. The machine learning model was able to use diverse data to identify groups of new policies with accident rates six times as great as the median; this grouping formed the basis of a new pricing strategy that improved profitability by more than 10 percent. Predictive analytics has been making an impact in medicine for several years now; an algorithm powered by deep learning won a Merck-sponsored design challenge to identify high-potential molecules for drug development.[98] More recently, a team from Houston Methodist developed an algorithm that translates text from the hospital's patient charts into a prediction of breast cancer

[97] Carlos A. Gomez-Uribe and Neil Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems*, volume 6, issue 4, January 2016.

[98] John Markoff, "Scientists see promise in deep-learning programs," *The New York Times*, November 23, 2012.

Industry applications of strategic optimization

Consumer: Optimize in-store product assortment to maximize sales

Pharmaceuticals: Optimize design of clinical trials, including patient selection

Telecom: Optimize capital expenditures across networks

Energy: Optimize blend of raw materials in refining or other processes

Public sector: Optimize resource allocation across various public infrastructure projects

Media: Identify features of optimal content and improve casting

Industry applications of real-time optimization capabilities

Consumer: Optimize logistics, procurement timing, and inventory distribution across warehouses and stores

Manufacturing: Optimize complex manufacturing processes in real time and determine where to dedicate resources to reduce bottlenecks and cycle time (particularly relevant in automotive and other types of advanced manufacturing)

Travel, transport, and logistics: Optimize routing in real time (airlines, logistics, and last-mile routing for delivery)

Energy: Optimize power plants based on energy pricing, weather, and other real-time data

risk 30 times as fast as a human can.[99] The opportunities are far-reaching, and the potential value is high. (See box on the preceding page for a summary of predictive analytics applications.)

A third set of uses relates to strategic optimization. This involves using insights from data and machine learning to make mid- to long-term decisions that improve organizational performance (see box at left).

There are numerous examples of how this could be applied in various industries. Formula One teams pour hundreds of millions of dollars annually into development, continually aiming for incremental technological improvements that can boost speed. Three F1 teams recently turned to machine learning to hold down costs in their aerodynamics operations divisions. Building on years of diverse project data, they looked for patterns that influenced the efficiency of a given engineering project and were able to achieve millions of dollars in savings. In insurance, a large auto insurer used machine learning to improve the policy pricing provided by traditional models. Machine learning algorithms identified broad mispricing across the company's portfolio, with 60 percent of it mispriced by 10 percent or more.

A final set of uses identified by our survey respondents relates to the real-time optimization of operations and logistics (see box at left). These applications involve heightening the efficiency of routes, machinery, and other processes or equipment. Machine learning could create dramatic efficiencies in these areas by predicting failures, identifying bottlenecks, and automating processes and decisions.

In the oil industry, self-learning simulation models can adjust parameters and controls based on real-time well data. A mid-sized oil field in Southeast Asia used this application and generated production improvements of $80 million to $100 million annually. DHL is using machine learning to optimize its complex global logistics operation. Tremendous potential for real-time optimization remains untapped for many companies in functions such as managing distribution networks, managing inventory, and timing procurement.

ADVANCES IN DEEP LEARNING HAVE THE POTENTIAL TO GREATLY EXPAND THE SCOPE OF AUTOMATION

We also examined the potential of machine learning to automate activities we currently pay people to do. Previous MGI research examined the potential to automate 2,000 work activities in every occupation in the economy.[100] For each work activity, we identified the level

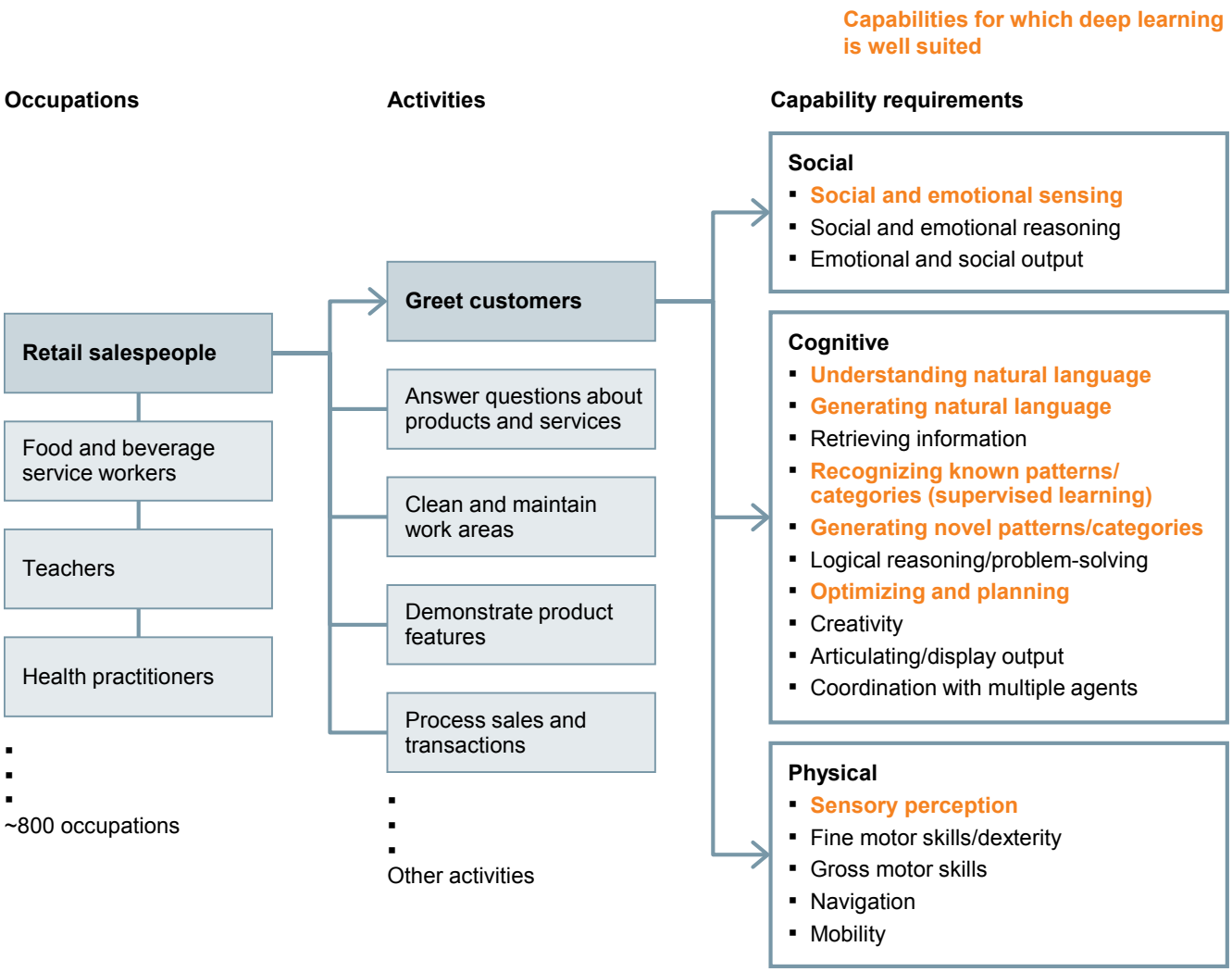
[99] Tejal A. Patel et al., "Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods: Natural Language Processing," *Cancer*, August 2016.

[100] These "detailed work activities" are defined by O*NET, a data collection program sponsored by the US Department of Labor. See Michael Chui, James Manyika, and Mehdi Miremadi, "Four fundamentals of workplace automation," *The McKinsey Quarterly*, November 2015.

of machine learning performance that would be required across the 18 different capabilities to successfully accomplish the activity (Exhibit 19).

Exhibit 19

Deep learning is well suited to develop seven out of 18 capabilities required in many work activities



NOTE: While this example illustrates the activities performed by a retail worker only, we analyzed some 2,000 activities across all occupations.

SOURCE: McKinsey Global Institute analysis

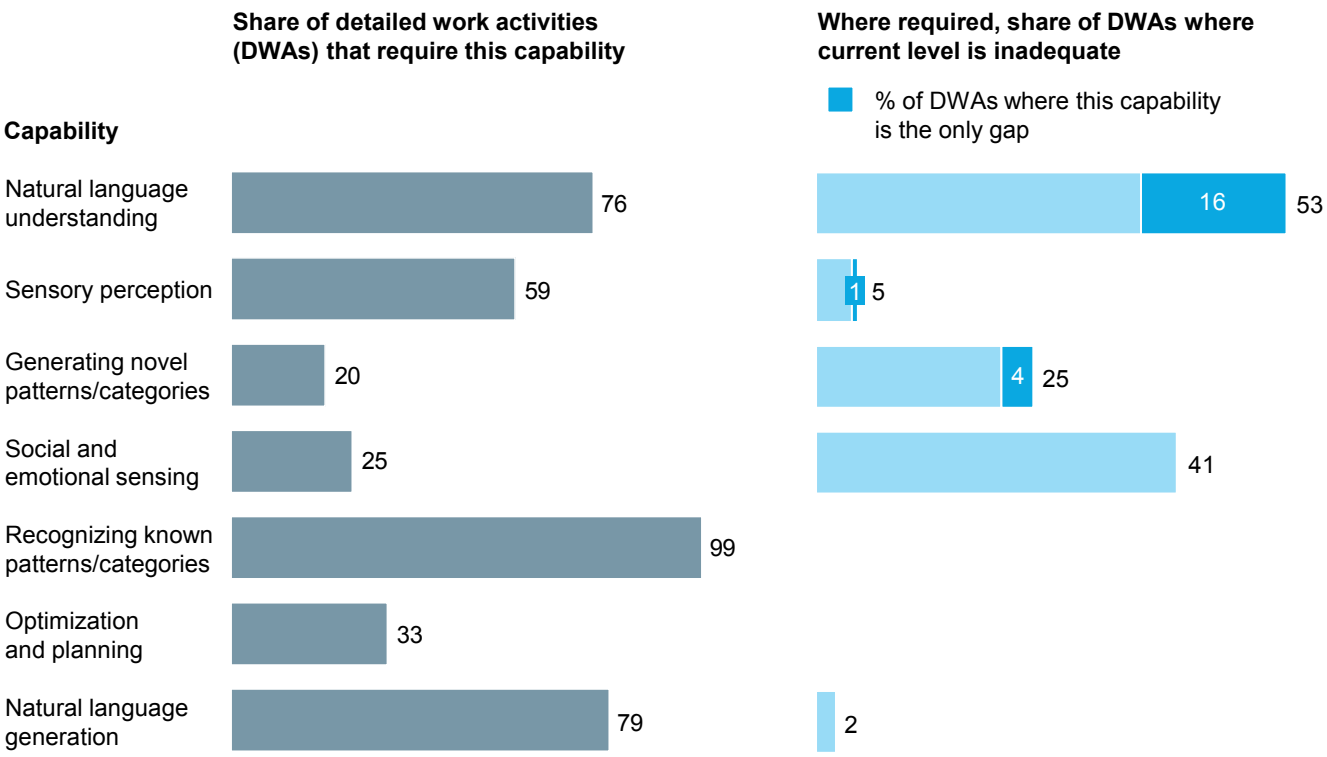
Seven of those 18 capabilities are well-suited to being implemented through the use of machine learning (Exhibit 20). The first striking observation is that almost all activities require capabilities that correlate with what machine learning can do. In fact, only four out of more than 2,000 detailed work activities (or 0.2 percent) do not require any of the seven machine learning capabilities. Recognizing known patterns, by itself, is needed in 99 percent of all activities to varying degrees; that is a fundamental capability of machine learning. Natural language generation, natural language understanding, and sensory perception are required for most work activities in the economy (79 percent, 76 percent, and 59 percent of all detailed work activities, respectively). This is not to say that such a high share of jobs is likely to be automated anytime soon, but it does underscore the wide applicability of machine learning in many workplaces.

For some of these capabilities, currently demonstrated technology has already attained sufficient performance to automate work activities. For instance, machine learning capabilities in pattern recognition generally already exceed median human performance. In media, artificial intelligence bots are already generating simple news and sports summary articles. One even wrote the screenplay for a movie titled *Sunspring*.[101] Improvements in natural language understanding will only expand the possible uses in many fields.

Exhibit 20

Improvements in natural learning understanding and generation as well as social sensing would have the biggest impact on expanding the number of work activities that deep learning could technically automate

%



SOURCE: McKinsey Global Institute analysis

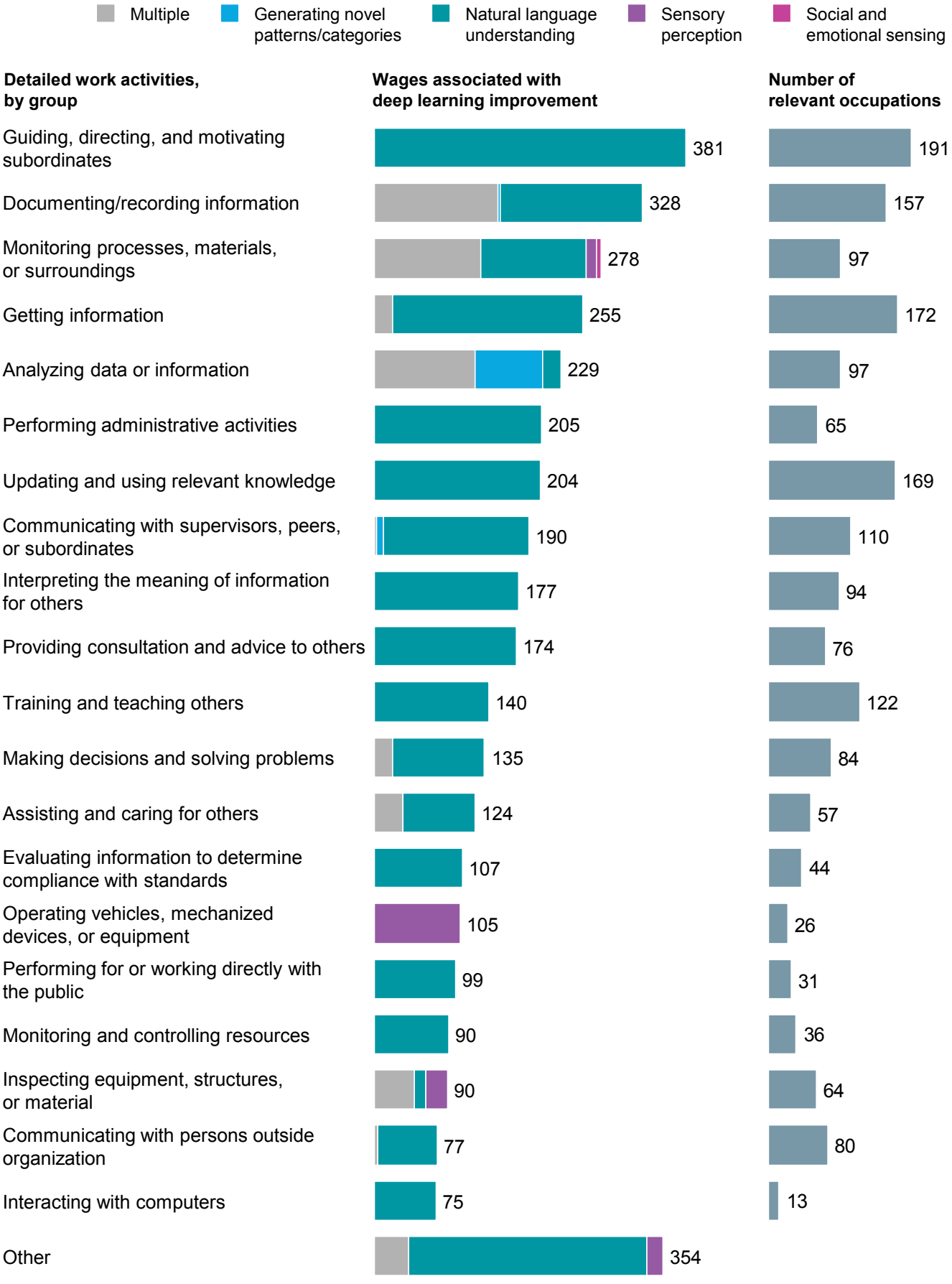
Previous MGI research on automation found that 45 percent of all work activities, associated with $14.6 trillion of wages globally, has the potential to be automated by adapting currently demonstrated technology.[102] Some 80 percent of that could be implemented by using existing machine learning capabilities. But deep learning is still in its early stages. Improvements in its capabilities, particularly in natural language understanding, suggest the potential to unleash an even greater degree of automation. In 16 percent of work activities that require the use of language, increasing the performance of machine learning in natural language understanding is the only barrier to being able to automate that work activity. Improving these capabilities alone could lead to an additional $3 trillion in wage impact (Exhibit 21). Advances in sensory perception and generating novel patterns and categories, which could be enabled by deep learning, could further increase the number of activities that could be automated.

[101] Jacob Brogan, "An artificial intelligence scripted this short film, but humans are still the real stars," *Slate*, June 9, 2016.

[102] Michael Chui, James Manyika, and Mehdi Miremadi, "Four fundamentals of workplace automation," *McKinsey Quarterly,* November 2015, and "Where machines could replace humans—and where they can't (yet)," *McKinsey Quarterly*, July 2016.

Exhibit 21

Top 20 groups of work activities by wages that could be affected by improved deep learning capabilities



SOURCE: McKinsey Global Institute analysis

To characterize the impact of these improvements, we assessed the occupations and wages at stake across a consolidated list of 37 work activity groups defined by the US Bureau of Labor Statistics. Exhibit 21, above, lists this information for the 20 groups with the highest potential wage impact associated with deep learning improvement. The most valuable activity groups are prevalent across a variety of occupations: guiding subordinates, documenting information, monitoring surroundings, and gathering information. For each of these activity groups, we also identified where improvements in specific deep learning capabilities could lead to automation. Most notably, improving natural language understanding would enable automation across all activity groups, with the greatest impact in basic management and supervisory activities (guiding, directing, and motivating subordinates), information gathering, and performing administrative activities. Sensory perception has particular impact in automating vehicles or other machinery, while generating novel patterns is relevant for data analyses.

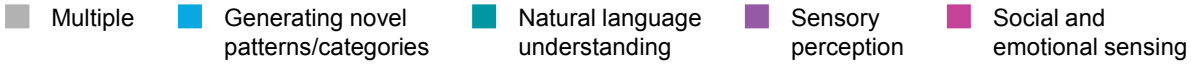
We also examined where the application of deep learning to automation could create the largest potential wage impact (Exhibit 22). Occupations like court reporters and interpreters could see nearly their entire jobs automated from improvements in the natural language understanding capabilities of machine learning. However, since these occupations employ relatively few people and pay low to medium wages, the expected overall impact will be small. Customer service representative is the lone occupation that makes the top 10 list for potential total wage impact and lends itself to automation across most of its work activities. The impact on a suite of frontline supervisory roles would also be large in dollar terms, as the primary activities for this group are guiding, directing, and motivating subordinates. Deep learning is likely to have a major impact on occupations with primarily administrative duties; these include executive assistants, cashiers, and waitstaff. Large numbers of people are employed in these occupations, which points to the possibility of significant job displacement. A significant percentage of the activities associated with a set of higher-paying jobs such as lawyers and nurses could also be automated with advances in machine learning.

While the potential of machine learning in general and deep learning in particular is exciting and wide-ranging, there are real concerns related to their development and potential deployment. Some of these were present even prior to the big data age, such as privacy, data security, and data ownership. But an additional set of new challenges has arisen.

First, deep learning has a drawback that poses a barrier to adoption in certain applications. The models that deep learning produces are opaque. As of today, it is relatively difficult to decipher how deep neural networks reach the insights and conclusions that they do. They are still a “black box.” However, researchers are working to create less opaque systems by doing the forensics to help people understand how these highly complex, trained models come to the conclusions that they do based on thousands or millions of connections and “weights” between nodes. For example, after AlphaGo played Lee Sedol, the world champion in the game of Go, the researchers who built that system were able to uncover what it was “thinking” when it made certain moves. But this is still a challenge, and when the mechanism behind a model is not understood, this can be potentially disqualifying in certain situations. Some decisions (such as hiring and granting loans) need to be transparent for legal reasons. Trying to run experiments or tweak variables can be difficult in a deep neural network; Google, for example, has hesitated until recently to use deep learning for its search algorithm for exactly this reason. There is also the matter of trust. It can be difficult for decision makers and customers to commit to insights that are generated in a non-transparent way, especially where those insights are counterintuitive. Medical use cases could fall into this category. This is not to say that model opacity will forever be a problem with deep neural networks, but for now, it must be noted that it can be barrier for adoption in certain use cases.

Exhibit 22

Improvements in deep learning (DL) could affect billions of dollars in wages in ten occupations globally



Occupations	% of time spent on activities that could be automated if DL improves (by DWA group)[1]	Most frequently performed group of DWAs that could be automated if DL improves	Global employ-ment Million	Hourly wage $	Global wages that DL could automate $ billion
Secretaries and administrative assistants, except legal, medical, and executive	28	Interacting with computers to enter data, process information, etc.	48.2	3.90	109.8
Business operations specialists, all other	30	Performing administrative activities	6.1	24.68	94.2
Managers, all other	27	Monitoring processes, materials, or surroundings	8.3	18.25	86.7
First-line supervisors of office and administrative support workers	35	Interpreting the meaning of information for others	12.8	8.75	81.5
Cashiers	18	Performing administrative activities	68.1	3.18	81.5
First-line supervisors of retail sales workers	13	Guiding, directing, and motivating subordinates	19.7	15.02	77.4
Industrial engineers	20	Getting information	8.0	20.60	69.4
Customer service representatives	51	Performing for or working directly with the public	6.9	9.35	67.4
Lawyers	31	Providing consultation and advice to others	2.3	41.14	61.8
First-line supervisors of helpers, laborers, and material movers	24	Organizing, planning, and prioritizing work	8.5	12.73	54.2

1 Detailed work activity. There are 37 total DWA groups.

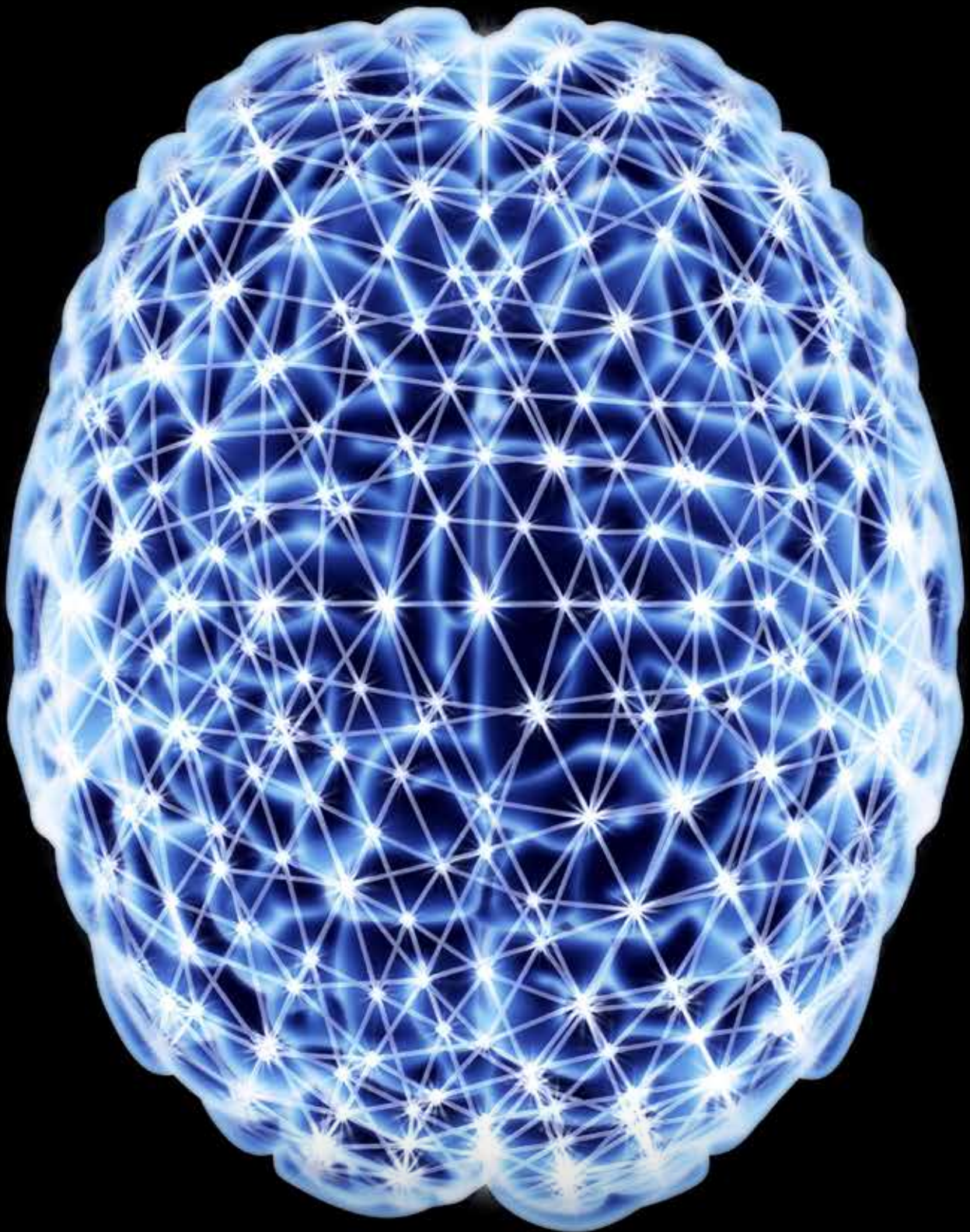
SOURCE: National labor and statistical sources; McKinsey Global Institute analysis

Second, there are ethical questions surrounding machine intelligence. A dystopian future in which machine superintelligence controls humanity has long been the stuff of science fiction, but there are more immediate issues to grapple with. One set of ethical concerns is related to real-world biases. Since the real world is racist, sexist, and biased in many other ways, real-world data that is fed into algorithms can also have these features—and when machine learning algorithms learn from biased training data, they internalize the biases, exacerbating those problems. A related concern revolves around deciding whose ethical guidelines will be encoded in the decision making of intelligence and who will be held responsible for the algorithm's conclusions. These questions have entered the public sphere most recently in the context of autonomous vehicles, and they seem likely to be relevant in many other contexts. There are additional concerns related to how intelligent automation could alter the nature of human interaction. Leading artificial intelligence experts, through such efforts as OpenAI and the Foundation for Responsible Robotics, have begun posing these questions—and more business leaders, policy makers, and thought leaders will need to shape the discourse.

Third, the potential risks of labor disruption from the use of deep learning to automate activities are becoming a critical debate, particularly in light of existing anxiety about the quantity and quality of available jobs. There is historical precedent for major shifts among sectors and changes in the nature of jobs. In the United States, the share of farm employment fell from 40 percent in 1900 to 2 percent in 2000; similarly, the share of manufacturing employment fell from roughly 25 percent in 1950 to less than 10 percent in 2010. In both circumstances, while some jobs disappeared, new ones were created, although what those new jobs would be could not be ascertained at the time. In 1950, few would have predicted that millions of people would be employed in information technology jobs in the following decades. But the past does not provide adequate assurances that sufficient numbers of new, quality jobs will be created at a sufficient rate. At the same time, many countries have or will soon have labor forces that are declining in size, requiring an acceleration of productivity to maintain historical rates of economic growth. A forthcoming MGI research report on automation will address the potential pace of automation in different economies. But certainly dealing with job displacement, retraining, and unemployment will require a complex interplay of government, private sector, and educational and training institutions, and it will be a significant debate and an ongoing challenge across society.

•••

The world could be on the cusp of untold disruption from machine learning today and deep learning in the near future. Both the opportunities and the risks are great. Organizations that are able to harness these capabilities effectively will be able to create significant value and differentiate themselves, while others will find themselves increasingly at a disadvantage.



© Science Photo Library — PASIEKA/Getty Images

TECHNICAL APPENDIX

This appendix provides details on the methodology employed in our research.

1. METHODOLOGY FOR ADOPTION RATES AND VALUE CAPTURE BY INDUSTRY

We considered the same industries we studied in our 2011 report: location-based services, US retail, US health care, the EU public sector, and manufacturing.[103]

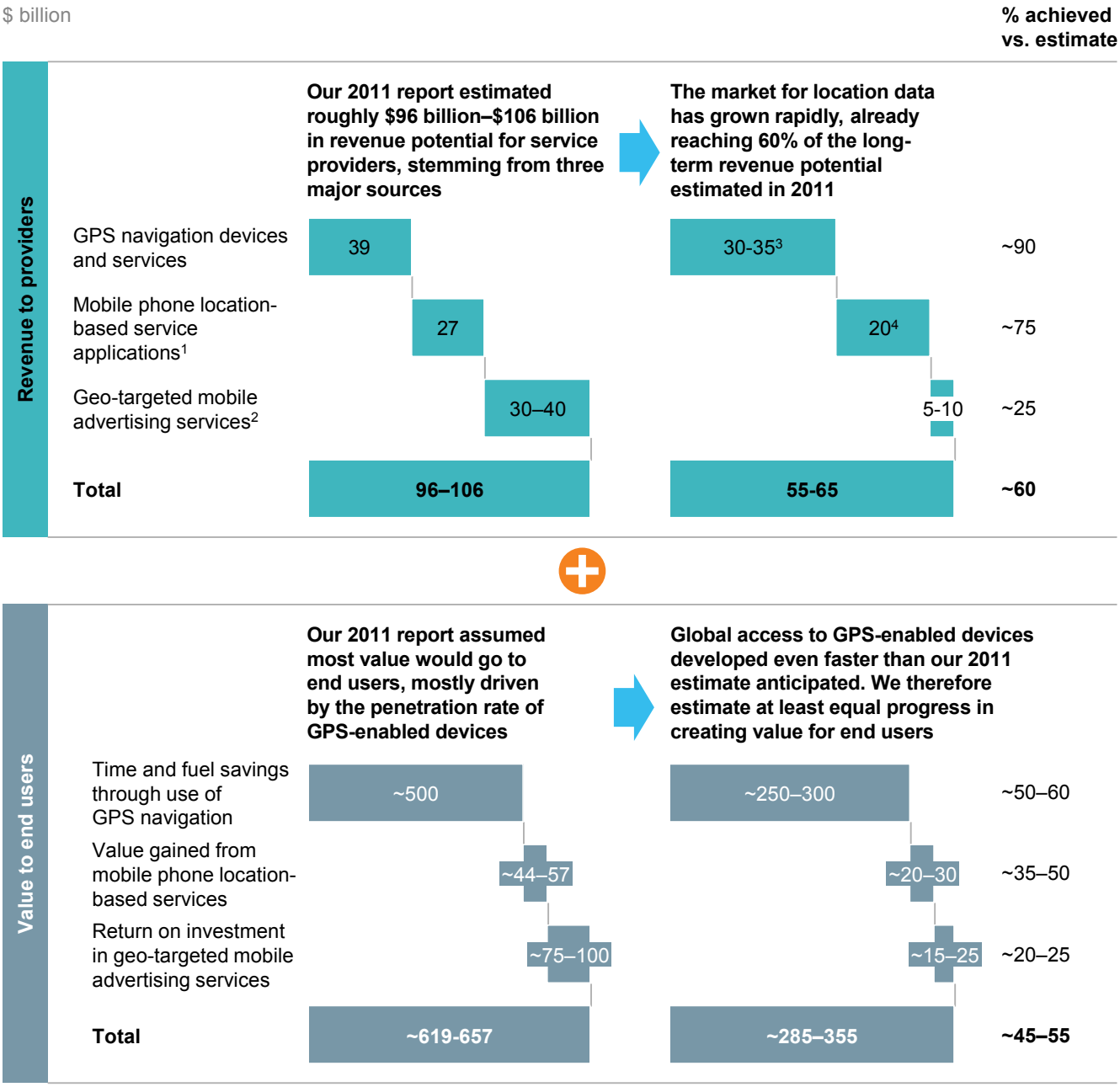
Location-based services

For revenue to service providers, we used the same methodology as described in our 2011 report to calculate revenue. We updated with numbers from 2015, which allowed us to calculate the improvement in revenue relative to the potential forecasted. We did this update only for the top three levers: GPS navigation devices and services, mobile phone location-based service applications, and geo-targeted mobile advertising services (Exhibit A1). These accounted for the vast majority of the value potential from the 2011 report. For value to end consumers, we again focused on the top three levers: time and fuel saving in traveling through access to GPS navigation, value gained from mobile phone location-based services, and return on investment in geo-targeted mobile advertising services. The forecasts we developed in 2011 were closely tied to mobile phone penetration rates. Given the current mobile phone penetration rate, we could interpolate the progress made against our expectation to arrive at the impact in these levers since 2011.

[103] *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, June 2011.

Exhibit A1

Location-based services have experienced strong growth, capturing a large share of the value potential we identified in 2011


$ billion
% achieved vs. estimate
Revenue to providers
Our 2011 report estimated roughly $96 billion–$106 billion in revenue potential for service providers, stemming from three major sources
The market for location data has grown rapidly, already reaching 60% of the long-term revenue potential estimated in 2011
GPS navigation devices and services
Mobile phone location-based service applications[1]
Geo-targeted mobile advertising services[2]
Total
39
27
30–40
96–106
30-35[3]
20[4]
5-10
55-65
~90
~75
~25
~60
Value to end users
Our 2011 report assumed most value would go to end users, mostly driven by the penetration rate of GPS-enabled devices
Global access to GPS-enabled devices developed even faster than our 2011 estimate anticipated. We therefore estimate at least equal progress in creating value for end users
Time and fuel savings through use of GPS navigation
Value gained from mobile phone location-based services
Return on investment in geo-targeted mobile advertising services
Total
~500
~44–57
~75–100
~619-657
~250–300
~20–30
~15–25
~285–355
~50–60
~35–50
~20–25
~45–55

1 Includes, for example, people tracking, location sharing, social community applications, city/regional guides/services, entertainment.
2 Derived as share of total mobile advertising market size.
3 PNDs and in-car devices/services.
4 Based on app stores; includes revenues generated from application stores (across operating systems).
NOTE: Not to scale. Numbers may not sum due to rounding.

SOURCE: 451 Research 2016; GNSS Market Report 2015; Marketsandmarkets 2016; Ovum 2015; Statista 2016; McKinsey Global Institute analysis

US retail, US health care, and the EU public sector

The assessments for US retail, US health care, and the EU public sector were carried out very similarly. For each lever identified in the 2011 report, we interviewed multiple industry experts to obtain data and perspective on the adoption progress since 2011 and on the value captured by companies that have adopted that lever. The current adoption level multiplied by the percentage of value capture by adopters gives us the industry's value capture. We then weighted each lever by the total value capture amount (in dollars) to

determine the aggregate impact. For each lever, we also searched for external surveys or data that could cross-validate the expert assumptions, and where possible we used this data to supplement interviews (Exhibits A2 to A4).

Exhibit A2

US retail has made solid progress in integrating data and analytics, especially among larger players, but further potential can be realized

Categories	Marketing and sales	Merchandising	Operations	Supply chain	New business models
Levers	▪ Improved cross-selling ▪ Location-based marketing ▪ Improved customer segmentation ▪ In-store behavior data ▪ Sentiment analysis ▪ Enhanced multi-channel closed-loop marketing	▪ Pricing optimization ▪ Assortment optimization ▪ Placement optimization	▪ Performance transparency ▪ Labor resource optimization	▪ Improved inventory management ▪ Improved sourcing ▪ Transportation optimization	▪ Price comparison services ▪ Data selling
Operating margin, 2011 % change	9–33	8–43	4–13	6–36	-4 to -28
Adoption rate, 2016 %	50–80	60–90	80–100	70–90	30–80
Average value realized by adopters %	30–60	25–50	20–50	50–70	20–50
Operating margin value realized % change	2–12	2–16	1–5	3–19	-1 to -10

30–40% value captured

SOURCE: McKinsey Global Institute analysis

Exhibit A3

US health care has seen increased adoption but has captured only ~10 percent of the potential value identified in MGI's 2011 research

Categories	Clinical operations	Accounting/ pricing	R&D	New business models	Public health
Levers	▪ Comparative effectiveness research ▪ Clinical decision support tools ▪ Transparency about medical data ▪ Remote patient monitoring ▪ Advanced analytics applied to patient profiles	▪ Automated systems ▪ Health economics and outcomes research and performance-based pricing plans	▪ Predictive modeling ▪ Statistical tools and algorithms to improve clinical trial design ▪ Analyzing clinical trials data ▪ Personalized medicine ▪ Analyzing disease patterns	▪ Aggregating and synthesizing patient clinical records and claims datasets ▪ Online platforms and communities	▪ Public health surveillance
Value potential, 2011 USD billions	164	47	105	5	9
Adoption rate, 2016 %	30–70	10–20	30–50	90–100	60–80
Average value realized by adopters %	10–30	20–30	30–50	40–100	60–80
Value potential realized USD billions	5–35	1–3	10–27	2–5	3–6

10–20% value captured

SOURCE: McKinsey Global Institute analysis

Exhibit A4

The EU public sector has captured less than 15 percent of the potential value estimated in MGI's 2011 research

	2011 report estimates		Current state of EU public-sector agencies	
Category	Assumed addressable baseline	Impact potential on baseline %	Adoption rate today compared to 2011 % change	Impact achieved by today's adopters %
Operational efficiency savings	20–25% of government operating expenditure net of transfer payment	-15–20	+10–20	0–50
Reduction in fraud and error	1–3% of government transfer payments	-30–40	+5–15	25–75
Increase in tax collection	Tax gap (estimated at 5–10% of tax revenue collected)	-10–20	+10–20	50–100

Effectively, less than 15% of the €150 billion–€300 billion in annual efficiencies estimated in 2011 likely has been achieved

NOTE: Similar observations hold true for governments around the world (e.g., United States closer to 20%; Israel closer to 10%).

SOURCE: McKinsey Global Institute analysis

Manufacturing

The manufacturing sector was assessed in a similar way through a combination of expert interviews and outside sources per lever. Since the levers from the 2011 report did not total to a dollar figure or margin in aggregate (each lever had its own measure for impact), we took the average value capture across the levers as the industry view (Exhibit A5).

Exhibit A5

Manufacturing has lagged in adoption, with most value being captured by a small group of advanced players

%

	Lever	Impact potential identified in 2011 report	Current state across manufacturing industries: Adoption rate today compared to 2011	Current state across manufacturing industries: Estimated impact achieved by today's adopters
R&D	▪ Concurrent engineering/ product lifecycle management	▪ -20–50 product development costs ▪ -20–50 time to market	+0–10	▪ -10–30 product cost
	▪ Design-to-value	▪ +30 gross margin	+15–20	▪ 10–40 gross margin
	▪ Crowd sourcing	▪ -25 product development costs	n/a	▪ n/a
Supply chain management	▪ Demand forecasting/ shaping and supply planning	▪ +2–3 profit margin ▪ -3-7 working capital (one time)	+5–10	▪ -10–15 product cost ▪ -5 working capital (one time)
Production	▪ "Digital factories" for lean manufacturing	▪ -10–25 operating costs ▪ Up to +7 revenue	+5–10	▪ -10–15 operating cost
	▪ Sensor data-driven operations analytics	▪ -10–50 assembly costs ▪ +2 revenue		
After-sales	▪ Product sensor data analysis for after-sales service	▪ -10–40 maintenance costs ▪ +10 annual production	+10–20	▪ -10–15 operating cost (for aftersales clients)

SOURCE: McKinsey Global Institute analysis

2. METHODOLOGY FOR TALENT SUPPLY AND DEMAND ANALYSIS

The supply and demand analysis of data scientists consists of three main pieces: current supply/demand "equilibrium," new supply, and future demand. As a starting point, we assume that the supply and demand are in equilibrium in 2014. All the estimates are for the United States, the only location for which we could obtain relevant data.

To estimate the number of data scientists in the United States in 2014, we proceed in the following fashion. First, we define the skills of data science (as a subset of skills defined in the Burning Glass database) to be: data science, machine learning, data mining, statistical modeling, predictive analytics, predictive models, natural language processing, logistic regression, support vector machines (SVM), neural networks, naive Bayes, k-means, principal component analysis, R or WEKA. Also, we did not include skills such as Python because Python is also listed as a skill for many other roles (e.g., software developer), and Hadoop, which we consider to be more characteristic of the roles of data architects and engineers.

Then, for each Standard Occupational Classification (SOC) occupation code in the United States, we use Burning Glass data to identify the share of job postings in each occupation that include data science skills (based on our definition of data science skills above). We then multiply this share by the Bureau of Labor Statistics (BLS) employment for each SOC

occupation for 2014 in its employment projections data. A total of 335 SOC occupations included these data science skills. The highest share of data scientists (44 percent) reside in SOC occupation 15-1111, which is computer and information research scientists.

To calculate the new supply of data scientists to 2024, we assume that data science jobs will require a bachelor's degree at a minimum. We estimate the share of graduates in each level (bachelor's or master's degree or PhD) and field of study who would have the capabilities for the work. We linearly project the new data scientist supply between 2014 and 2024 based on historical Integrated Postsecondary Education Data System (IPEDS) graduation figures. We net out people who might have more than one degree, the population of foreign students who likely will not stay in the United States, and people who leave the workforce due to retirement (albeit the average age of a data scientist is young enough that the number of data scientists retiring within a decade will likely be small).

To calculate demand for data scientists to 2024, we estimate the share of employees who are data scientists at companies that are at the "talent frontier," which we define as the largest companies in each sector that have the greatest share of data talent as a percentage of their employee base. So we first identify the ten largest companies (by market capitalization) in each sector in the United States. For these companies, we count the number of employees with "data scientist" and "data analyst" titles in LinkedIn's database and divide that number by the company's total employment count in LinkedIn. For each sector, we take a weighted average of the three largest companies with the highest talent share, as a proxy for the data talent frontier of the sector. Then we combine these shares with BLS employment projections by occupation for 2024 under two scenarios. In the low scenario, we assume that in each sector, the data talent share for the average company in 2024 will look like its sector frontier today. In the high scenario, we assume that the data talent share for the average information and communications technology (ICT), professional services, and health care and social assistance organization in 2024 will look like the frontier in ICT today. And we assume that the average company in any other sector in 2024 will look like the frontier in wholesale today.

To calculate new demand for business translators to 2024, we assumed a ratio of business translators to data scientists of 4:1 to 8:1. Since we project approximately 500,000 new data science jobs to 2024, we estimate two million to four million new business translator jobs to 2024. Linearly projecting Integrated Postsecondary Education Data System degree completion data, we expect between nine million and ten million STEM plus business graduates between now and 2024.

3. METHODOLOGY FOR DISRUPTIVE ARCHETYPE ANALYSIS

We identified six archetypes for data and analytics–enabled disruption in Chapter 4: business models enabled by orthogonal data, hyperscale platforms with real-time matching, radical personalization, massive data integration capabilities, data-driven discovery, and enhanced decision making. Three of these discussions include sizing estimates, and the methodology for each is described below.

Hyperscale real-time supply/demand matching

To size the economic impact from scaling hyperscale real-time supply/demand matching in mobility services, we drew on significant McKinsey expertise in automotive and mobility, including previous McKinsey research reports on IoT and the automotive industry.[104]

[104] *The internet of things: Mapping the value beyond the hype*, McKinsey Global Institute, June 2015, and *Automotive revolution—perspective towards 2030*, McKinsey Advanced Industries Practice, January 2016.

We focused on three primary categories to structure our approach: cost savings for consumers (vehicle purchasers/drivers), time savings for consumers, and benefits for the broader public.

Across these three categories, we identified seven individual levers. To size the economic impact for each lever, we estimated the value pool at stake for developed and developing regions separately and assumed a 10 to 30 percent capture rate by 2025–30. The value pool estimates were constructed as follows:

- **Fewer vehicle purchases** (cost savings): We estimated the total savings from consumers switching to mobility services where economically advantageous. Specifically, we estimated the cost savings associated with switching at different annual mileage totals and aggregated across the proportion of consumers falling in each bucket. For developing regions, we adjusted for the lower cost of ownership and lower cost of transportation but otherwise adopted the same methodology.

- **Reduced parking cost** (cost savings): We estimated the total savings for consumers from lower demand for parking spots. Focusing on commercial parking, we combined the share of vehicles, the average price for parking, and the total number of working hours to determine the total value pool.

- **Time savings from parking** (time savings): We estimated the total time spent searching for parking by urban vehicles, based on a lower estimate of 15 percent of total time in vehicles and the share of cars in cities. Applying the time value for developed and developing regions yields the total value at stake.

- **Reduced fuel cost from parking** (cost savings): Based on the time savings implied by the previous lever, we estimated the associated fuel value by applying the estimated fuel efficiency and fuel cost for developed and developing regions.

- **Reduced pollution from parking** (public benefit): Based on the time and fuel savings implied by the previous two levers, we estimate the carbon dioxide emissions from parking searches. Applying the value per ton of CO_2 produces the savings estimates.

- **Productive time during rides** (time savings): We estimated the total time spent driving by individuals in both developed and developing regions and assumed 50 percent of this time could be spent productively. Applying the time value for these segments produced the total value pool.

- **Reduced number of accidents** (public benefit): We estimated the number of accidents in both developed and developing regions based on World Health Organization (WHO) figures from 2014. We leveraged estimates from recent studies on potential reduction in accident rates from mobility services, combined with estimated costs per accident, to determine associated savings.

Radical personalization

For radical personalization, we estimated the economic impact across six levers:

- **Improve treatment efficacy by assessing health risks and adherence through monitoring:** This lever was sized using the same assumptions made in our IoT report. We looked at the disability-adjusted life years (DALY) impact from IoT technologies for each disease area and sized the total across countries. We updated the DALY numbers to 2016 numbers, obtained from the WHO. The assumed impact $/DALY we used for this report was the WHO recommendation of three times GDP per capita.

- **Improve overall wellness through tailored programs:** This lever was sized using the same assumptions made in our IoT report. No changes were made except for the assumed impact $/DALY, using the WHO recommendation of three times GDP per capita.

- **Proactive risk assessment to anticipate complications and reduce hospital care:** We used estimates for unnecessary hospital care from the United States and the United Kingdom to size the percentage of total hospital care that could be avoided. Scaled across all countries (and applied to hospital expenditures only), we developed an estimate for all high-income countries. Given the sparsity of health insurance and lower overall health-care spend in lower and middle income countries, we sized the potential for cost savings from only the United States and other high-income countries.

- **Direct patients to the highest-value care to reduce costs (e.g., pricing transparency):** Several studies showed the impact of pricing transparency on consumer choice and overall cost savings. Given these percentages, we developed low and high estimates for impact from pricing transparency. We then applied these impacts to the share of national health-care costs paid by private insurers (whose rates will vary by physician) to arrive at our estimate for impact. Given the sparsity of health insurance and lower overall health-care spend in lower- and middle-income countries, we sized the potential for cost savings from only the United States and other high-income countries.

- **Deliver the right treatment and optimize the dose:** We based this sizing on the methodology and assumptions used in an analysis of genomics in a 2013 MGI report.[105] Genomics is a reasonable proxy for precision medicine and treatment selection. In the 2013 report, we sized the impact in cancer, diabetes, and cardiovascular disease. In this report, we used similar estimates across all non-communicable diseases and estimated the total DALY impact. Again, our assumed impact $/DALY we used for this report was the WHO recommendation of three times GDP per capita.

- **Eliminate treatments that are ineffective for that individual:** We followed the same methodology as in lever 5 to determine the applicable population for these treatments. Using US numbers for cost per disease area, we determined the total costs at stake in these treatment areas for precision medicine. We then assumed that some costs could be avoided due to precision screening, based on our earlier estimates of the cost savings possible from companion diagnostics in our 2011 report (lever for personalized medicine). This gave us our total impact in cost savings.

[105] *Disruptive technologies: Advances that will transform life, business, and the global economy*, McKinsey Global Institute, May 2013.

Massive data integration

In analyzing the value creation potential of improved data integration in retail banking, we identified 20 levers that have an impact on the revenue and cost of a sample retail bank. We group the levers into three categories depending on how much of the revenue and cost impact comes from data and analytics and how much from digital. On one end, in pure data and analytics levers, we assign the levers where advanced analytics and a considerable amount of data are needed, e.g., next product to buy recommendations and dynamic pricing. On the other end, in pure digital levers, we include levers such as reducing the physical footprint of branches and ATMs as a result of migrating customer to digital channels. The third category in between consists of levers where both data and analytics and digital are needed for value creation, such as efficiency gains from automating business support functions where both digitalization of the processes and automated decision making by analytics are needed.

To convert the single bank view into global value creation potential, we first leave out the revenue levers that are shifting value among players and then convert the remaining value-creating revenue levers into productivity gains that impact cost. For cost baseline numbers of the industry, we use the McKinsey Panorama global banking database to obtain the baseline separately for developed and emerging markets (2015 estimates for interest revenue, fee revenue and risk cost; 2014 cost-to-income ratio estimates for converting the revenue into operating cost). These baseline numbers are then multiplied with the efficiency potential associated with a single bank to create a view of the overall potential for the industry.

4. METHODOLOGY FOR DETERMINING THE AUTOMATION POTENTIAL OF MACHINE LEARNING

Here, we discuss our methodology on identifying use cases of machine learning, sizing those opportunities, and restricting them based on data availability.

To determine where machine learning is most relevant, we began by collecting as many possible use cases as we could across 12 industry groupings. This use-case driven approach ensured that our perspective on machine learning was grounded in actual industry-specific applications and could be categorized or grouped across industries.

To collect these use cases, we first characterized machine learning capabilities or problem types. We used this list of problem types to co-create a list of more than 300 use cases with about 50 industry experts. The use cases were also classified across nine “use case types” as recurring themes emerged in our analysis. This grouping is used in a number of exhibits in this appendix.

The 300 use cases formed the basis for our analysis of the potential of machine learning. Across all of these use cases, we constructed two separate indexes to characterize the scope and scale.

First, we estimated the value potential by conducting a survey across internal McKinsey experts aligned to the industries included in our report. In this survey, respondents were asked to rank the top three use cases in their industry out of a prioritized list of ten. Then, for each of the 120 use cases we constructed a value index by weighting the number of times each use case was ranked first by three respondents, second by two, and third by one.

The landscape of impact scoring across use case types and industry groups is presented below (Exhibit A6). We observe that there is broad potential for different use case types across industries, with a few standouts. In particular, radical personalization in media and predictive maintenance in energy appear to be higher impact. Predictive analytics appears to have value across industries.

Exhibit A6

Machine learning has great impact potential across industries and use case types

Problem type	Automotive	Manufacturing	Consumer	Finance	Agriculture	Energy	Health care	Pharma-ceuticals	Public/ social	Media	Telecom	Transport and logistics
Real-time optimization												
Strategic optimization												
Predictive analytics												
Predictive maintenance												
Radical personalization												
Discover new trends/anomalies												
Forecasting												
Process unstructured data												

SOURCE: McKinsey Global Institute analysis

Second, we assessed the data richness available for each of the 300 total use cases. The two criteria we focused on were as follows:

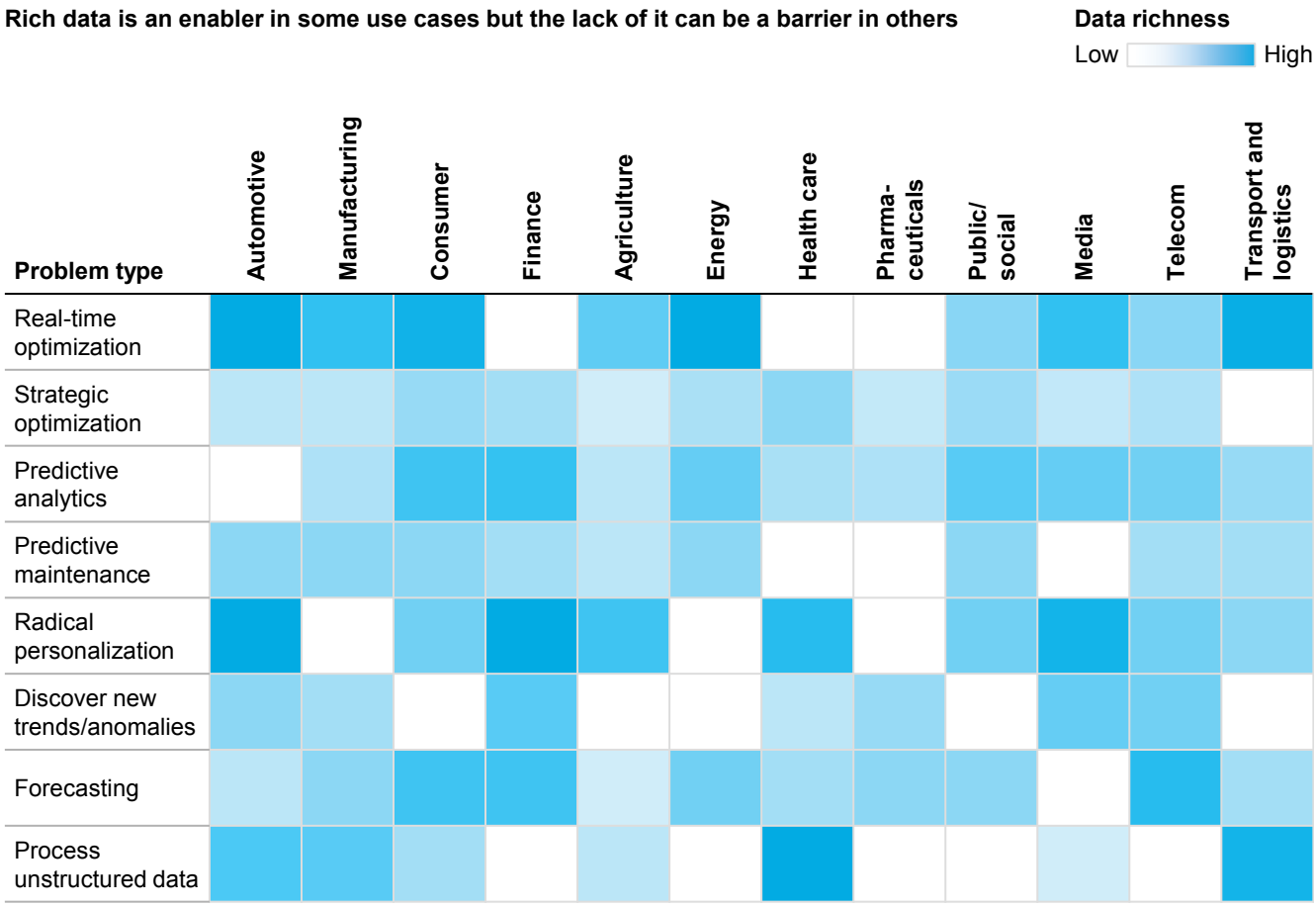
- **Volume of data:** Higher volumes of data mean more data points for the algorithms to learn from and thus, fewer barriers to capture the use case. So, the more data there are, the more beneficial it is for deep learning. This was taken as both a breadth and a frequency measure: how many individual data points there were as well as how often an interaction happened with each data point over the course of a year. After that, we multiplied the volume and frequency to arrive at an aggregate volume measure.

- **Variety of data:** In addition to volume, the variety of data can play a crucial role. The greater the variety, the easier it is for algorithms to learn. The different types of data that we scored were transactions (virtual transactions, including ones in the conventional sense as well as other key metrics such as web page visits), social, audio/video/images, IoT/sensor data, scientific/engineering data (from experiments/simulations), mobile/location (actual location-based data from mobile devices) and archives (historical records, mostly physical). Each use case was assigned a value based on the number of data varieties available in that context.

We aggregated both of these measures to form a proxy for data richness (Exhibit A7).

Exhibit A7

Rich data is an enabler in some use cases but the lack of it can be a barrier in others



SOURCE: McKinsey Global Institute analysis

In the exhibits that follow (A8 to A19), we have consolidated results from the impact and data richness scoring. For each industry group, we list the top ten use cases along with the impact score and data richness score. Impact scoring ranges from 0 to 3, while data richness ranges from 0 to 2 based on an average across breadth, frequency, and variety.

Exhibit A8

Machine learning opportunities in automotive

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Identify and navigate roads and obstructions in real-time for autonomous driving	Process unstructured data	1.5	2.0
Predict failure and recommend proactive maintenance on vehicle components	Predictive maintenance	0.9	1.0
Optimize manufacturing process in real time—determine where to dedicate resources to reduce bottlenecks and cycle time	Operations/logistics optimization (real time)	0.7	1.0
Optimize lane choices and path routing based on multi-modal data to reduce length of trip	Operations/logistics optimization (real time)	0.6	1.7
Recognize complex voice commands to access greater variety of services and features	Process unstructured data	0.5	0.3
Predict outcomes using fewer experiments to reduce experimental R&D costs (e.g., component testing, track testing)	Forecasting	0.4	0.3
Discover anomalies across fleet of vehicle sensor data to identify potential failure risks and pre-empt recalls	Discover new trends/anomalies	0.4	1.0
Optimize aggregate marketing mix and marketing spend	Price and product optimization	0.3	0.3
Personalize in-vehicle recommendations based on location data and passenger preferences (e.g., recommending local attractions or shops for proactive maintenance)	Radical personalization	0.3	1.7
Predict market potential for new product innovation	Forecasting	0.3	0.3

SOURCE: McKinsey Global Institute analysis

Exhibit A9

Machine learning opportunities in manufacturing

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Predict failure and recommend proactive maintenance for production and moving equipment	Predictive maintenance	1.3	1.0
Optimize complex manufacturing process in real time—determine where to dedicate resources to reduce bottlenecks and cycle time	Operations/logistics optimization (real time)	1.1	1.0
Predict future demand trends and potential constraints in supply chain	Forecasting	0.8	0.7
Identify design problems in pre-production to reduce ramp-up time to maximum output (i.e., yield ramp)	Predictive analytics	0.6	0.3
Identify root causes for low product yield (e.g., tool-/die-specific issues) in manufacturing	Discover new trends/ anomalies	0.5	0.7
Detect defects and quality issues during production using visual and other data	Process unstructured data	0.4	0.7
Optimize resource allocation in R&D and manufacturing, leveraging diverse data (e.g., communications, documentation) to track progress	Resource allocation	0.4	0.3
Optimize R&D experimental efficiency through process/operations	Operations/logistics optimization (real time)	0.4	0.3
Determine root causes for quality issues developed outside of manufacturing (e.g., during delivery, in supply chain)	Discover new trends/ anomalies	0.3	0.7
Identify critical factors to reduce number of required experiments for R&D and testing (e.g., component testing)	Predictive analytics	0.2	0.7

SOURCE: McKinsey Global Institute analysis

Exhibit A10

Machine learning opportunities in consumer products

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Optimize in-store product assortment to maximize sales	Price and product optimization	1.3	1.0
Personalize product recommendations and advertising to target individual consumers	Radical personalization	1.1	1.3
Optimize logistics, procurement timing, and inventory distribution across warehouses/stores	Operations/logistics optimization (real time)	1.1	1.7
Predict hyperregional sales/demand trends using real-time data	Forecasting	0.9	1.3
Predict market potential for new product innovation	Price and product optimization	0.6	0.3
Optimize aggregate marketing mix and marketing spend	Price and product optimization	0.5	0.3
Classify visual features from in-store video to collect data on product placement, product touches, and other key performance indicators for use in auditing or market research	Process unstructured data	0.2	0.7
Predict risk of individual employee churn and recommend solutions	Predictive maintenance	0.1	1.0
Route call-center cases based on multi-modal data (e.g., customer preferences, audio data) to increase customer satisfaction and reduce handling costs	Predictive analytics	0.1	1.3
Write product descriptions and ads for diverse product portfolio	Price and product optimization	0	0.3

SOURCE: McKinsey Global Institute analysis

Exhibit A11

Machine learning opportunities in finance

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Personalize product offerings to target individual consumers based on multi-modal data (mobile, social media, location, etc.)	Radical personalization	1.2	1.7
Identify fraudulent activity using customer transactions and other relevant data	Discover new trends/ anomalies	1.0	1.3
Evaluate customer credit risk using application and other relevant data for less biased real-time underwriting decisions	Predictive analytics	0.9	1.0
Predict risk of churn for individual customers/clients and recommend renegotiation strategy	Predictive maintenance	0.7	0.7
Discover new complex interactions in the financial system to support better risk modeling and stress testing	Discover new trends/ anomalies	0.7	0.7
Predict risk of loan delinquency and recommend proactive maintenance strategies	Predictive analytics	0.5	1.0
Predict asset price movements based on greater quantities of data (e.g., social media, video feeds) to inform trading strategies	Forecasting	0.4	1.3
Optimize labor staffing and distribution to reduce operational costs in front and back office	Resource allocation	0.4	0.7
Route call-center cases based on multi-modal data (e.g., customer preferences, audio data) to increase customer satisfaction and reduce handling costs	Predictive analytics	0.1	1.7
Optimize branch/ATM network based on diverse signals of demand (e.g., social data, transactions)	Resource allocation	0.1	0.3

SOURCE: McKinsey Global Institute analysis

Exhibit A12

Machine learning opportunities in agriculture

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Customize growing techniques specific to individual plot characteristics and relevant real-time data	Radical personalization	1.1	1.3
Optimize pricing in real time based on future market, weather, and other forecasts	Price and product optimization	1.0	0.7
Predict yield for farming or production leveraging IoT sensor data and other relevant data	Forecasting	0.8	0
Predict real-world results from fewer experiments to reduce experimental R&D costs (e.g., new crop testing)	Price and product optimization	0.8	0.3
Predict new high-value crop strains based on past crops, weather/soil trends, and other data	Price and product optimization	0.8	0.3
Predict product demand trends to inform production decisions	Forecasting	0.7	0
Optimize production process in real time—determine where to dedicate resources to reduce bottlenecks and cycle time	Operations/logistics optimization (real time)	0.5	0.7
Predict failure and recommend proactive maintenance for farming and production equipment	Predictive maintenance	0.3	0.3
Construct detailed map of farm characteristics based on aerial video	Process unstructured data	0.1	0.3
Optimize purchasing mix across suppliers and locations	Resource allocation	0.1	0

SOURCE: McKinsey Global Institute analysis

Exhibit A13

Machine learning opportunities in energy

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Predict failure and recommend proactive maintenance for mining, drilling, power generation, and moving equipment	Predictive maintenance	1.6	0.7
Replicate human-made decisions in control room environments to reduce cost and human error	Predictive analytics	0.9	1.3
Optimize energy scheduling/dispatch of power plants based on energy pricing, weather, and other real-time data	Operations/logistics optimization (real time)	0.9	1.3
Optimize blend and timing of raw materials in refining and similar processes	Resource allocation	0.7	0.3
Optimize mine plans based on drilling samples, past sites, and other data	Resource allocation	0.5	0.7
Predict energy demand trends based on multi-modal data	Forecasting	0.5	1.7
Optimize aggregate pricing and promotional targeting to energy customers	Price and product optimization	0.4	1.0
Interpolate ground composition to reduce necessary exploratory drilling samples	Forecasting	0.2	0.3
Predict lifetime value and risk of churn for individual customers	Predictive analytics	0.2	1.0
Optimize specifications in construction for power generation equipment based on previous sites and other relevant data	Resource allocation	0.1	0.7

SOURCE: McKinsey Global Institute analysis

Exhibit A14

Machine learning opportunities in health care

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Diagnose known diseases from scans, biopsies, audio, and other data	Predictive analytics	1.4	0.3
Predict personalized health outcomes to optimize recommended treatment	Radical personalization	1.2	1.3
Optimize labor staffing and resource allocation to reduce bottlenecks	Resource allocation	0.7	0.7
Identify fraud, waste, and abuse patterns in diverse clinical and operations data	Discover new trends/ anomalies	0.6	0.3
Predict individual hospital admission rates using historical and real-time data	Forecasting	0.5	0.7
Triage patient cases during hospital admission using patient data, audio, and video	Predictive analytics	0.5	0.3
Personalize messaging and approach (e.g., nudges) to improve wellness and adherence	Radical personalization	0.4	1.3
Evaluate doctor performance and provide outcome-improving feedback	Predictive analytics	0.3	0.3
Detect major trauma events from wearables sensor data and signal emergency response	Process unstructured data	0.2	1.7
Predict lifetime value and risk of churn for individual customers	Predictive analytics	0.2	1.0

SOURCE: McKinsey Global Institute analysis

Exhibit A15

Machine learning opportunities in pharmaceuticals

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Optimize design of clinical trials, including label writing and patient selection	Price and product optimization	1.3	0
Predict outcomes from fewer or diverse (e.g., animal testing) experiments to reduce experimental R&D costs and time to market	Predictive analytics	0.9	0.3
Predict risk of individual patient churn and optimal corrective strategy to maintain adherence	Predictive analytics	0.8	0.7
Identify target patient subgroups that are underserved (e.g., not diagnosed), and recommend mitigation strategy	Discover new trends/anomalies	0.7	1.3
Optimize resource allocation (e.g., reduce search space, prioritize drugs) in drug development using disease trends and other data	Resource allocation	0.7	0
Identify high-value providers and target marketing/product mix	Price and product optimization	0.7	0.7
Optimize product launch strategy based on past launches and relevant data	Price and product optimization	0.4	0
Discover new alternative applications for developed drugs (i.e., label expansion)	Discover new trends/ anomalies	0.3	0.3
Predict hyperregional product demand and relevant health trends	Forecasting	0.2	1.0
Optimize pricing strategy for drug portfolio	Price and product optimization	0.2	0

SOURCE: McKinsey Global Institute analysis

Exhibit A16

Machine learning opportunities in public and social sector

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Optimize public resource allocation for urban development to improve quality of life (e.g., reduce traffic, minimize pollution)	Resource allocation	1.1	0.7
Optimize public policy decisions (e.g., housing) to take into account greater set of complex interactions	Predictive analytics	0.9	1.0
Personalize public services to target individual citizens based on multi-modal data (mobile, social media, location, etc.)	Radical personalization	0.8	1.3
Replicate back-office decision processes for applications, permitting, tax auditing, etc.	Predictive analytics	0.6	1.0
Optimize procurement strategy to reduce costs for large government agencies (e.g., defense)	Resource allocation	0.5	0.3
Forecast macroeconomic variables based on vast government-proprietary and public data	Forecasting	0.5	0.7
Predict individualized educational and career paths to maximize engagement and success	Radical personalization	0.4	1.3
Predict risk of failure for physical assets (e.g., military, infrastructure) and recommend proactive maintenance	Predictive maintenance	0.3	1.0
Optimize labor allocation for publicly provided services to match demand	Resource allocation	0.3	1.3
Optimize pricing for government provided goods and services (e.g., tolls, park entrance fees)	Price and product optimization	0.2	0.3
Predict risk of illicit activity or terrorism using historical crime data, intelligence data, and other available sources (e.g., predictive policing)	Predictive analytics	0.2	1.7

SOURCE: McKinsey Global Institute analysis

Exhibit A17

Machine learning opportunities in media

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Personalize advertising and recommendations to target individual consumers based on multi-modal data (mobile, social media, location, etc.)	Radical personalization	1.9	1.3
Discover new trends in consumption patterns (e.g., viral content)	Discover new trends/ anomalies	1.2	1.3
Optimize pricing for services/offerings based on customer-specific data	Price and product optimization	0.7	1.0
Predict viewership for new content to optimize content production decisions using multi-modal data (mobile, social media, past productions, etc.)	Predictive analytics	0.7	1.0
Predict risk of individual customer churn based on multimodal data	Predictive analytics	0.7	1.3
Optimize aggregate marketing mix and marketing spend	Price and product optimization	0.3	0.3
Identify relevant features (e.g., copyright infringement, audience suitability) in media content	Process unstructured data	0.2	0
Identify high-value leads by combining internal and external data (press releases, etc.) for B2B customers	Discover new trends/ anomalies	0.2	1.0
Optimize resource allocation in network vs. current and future loads	Resource allocation	0.1	0
Optimize release dates and regional targeting for film rollouts	Price and product optimization	0.1	0.3

SOURCE: McKinsey Global Institute analysis

Exhibit A18

Machine learning opportunities in telecom

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Predict lifetime value and risk of churn for individual customers	Predictive analytics	1.3	1.3
Optimize capex investment across network	Resource allocation	1.1	0.3
Personalize strategy to target individual consumers based on multi-modal data (mobile, social media, location, etc.)	Radical personalization	1.1	1.3
Optimize call-center routing for individual calls (fewer agent-handled calls)	Predictive analytics	0.8	1.3
Discover new trends in consumer behavior using mobile data and other relevant data	Discover new trends/ anomalies	0.5	1.3
Predict failure and recommend proactive maintenance for fixed (substations, poles) and moving equipment	Predictive maintenance	0.4	0.7
Optimize micro-campaigns and short-term promotions	Price and product optimization	0.4	0.3
Predict regional demand trends for voice/data/other traffic	Forecasting	0.3	1.7
Replicate financial planning and other costly back-office functions	Predictive analytics	0.1	0.7
Optimize field-force labor allocation	Resource allocation	0.1	0.7

SOURCE: McKinsey Global Institute analysis

Exhibit A19

Machine learning opportunities in transport, travel, and logistics

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Optimize pricing and scheduling based on real-time demand updates (e.g., airlines, less than truckload shipping, mobility services)	Price and product optimization	1.7	1.3
Predict failure and recommend proactive maintenance for planes, trucks, and other moving equipment	Predictive maintenance	1.1	0.7
Optimize routing in real time (e.g., airlines, logistics, last mile routing for complex event processing)	Operations/logistics optimization (real time)	0.9	1.3
Optimize staffing levels and asset placement in real time	Operations/logistics optimization (real time)	0.5	0.3
Personalize loyalty programs and promotional offerings to individual customers	Radical personalization	0.4	1.0
Personalize product recommendations to target individual consumers	Radical personalization	0.4	1.0
Predict hyperregional sales/demand trends	Forecasting	0.3	0.7
Identify performance and risk for drivers/pilots through driving patterns and other data	Predictive analytics	0.2	1.0
Predict lifetime value and risk of churn for individual customers	Predictive analytics	0.2	0.3
Read addresses/bar codes in mail/parcel sorting machines to improve efficiency and reduce human error	Process unstructured data	0.2	1.3

SOURCE: McKinsey Global Institute analysis

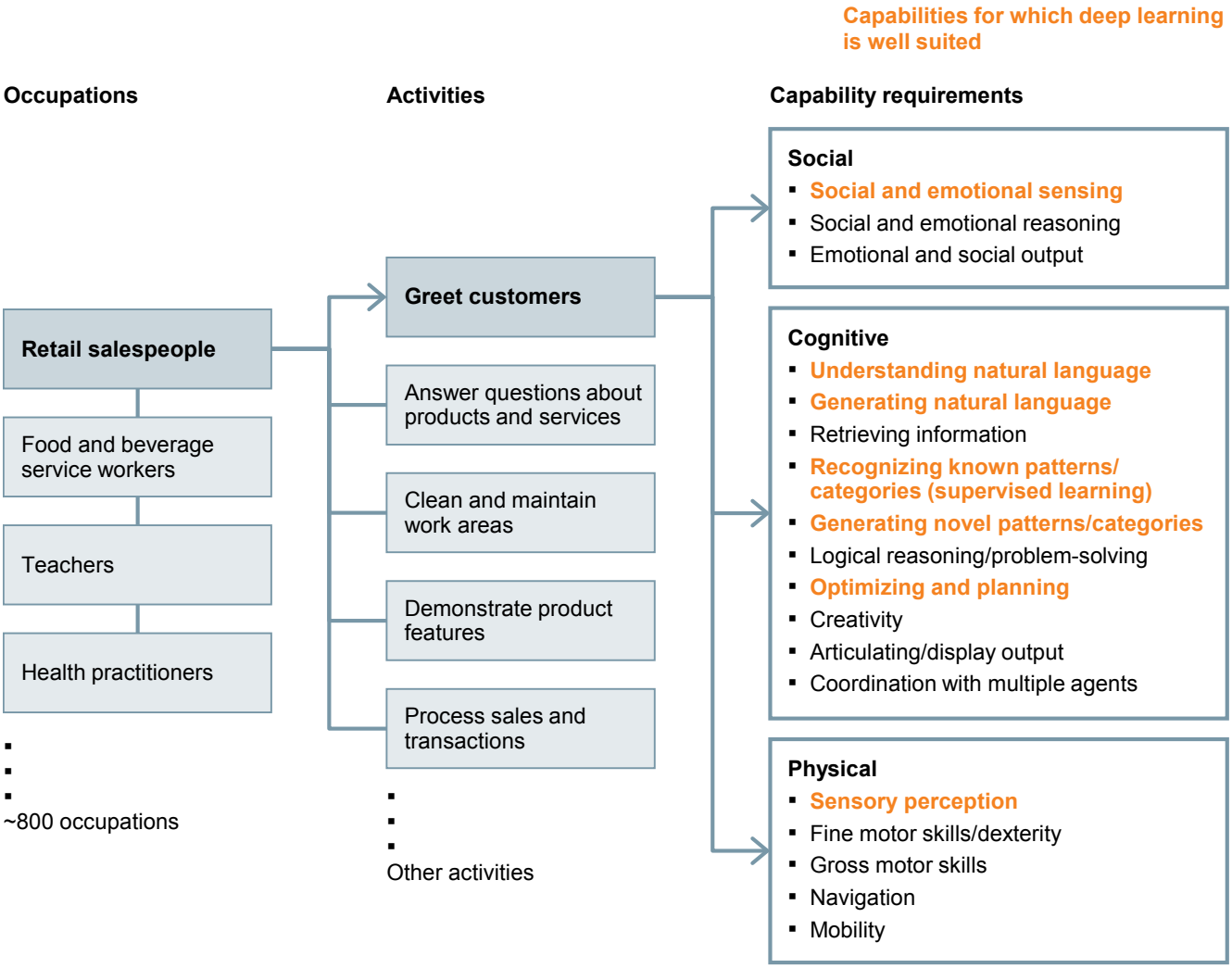
Detailed work activities

To determine which work activities could be performed by or affected by deep learning, we rely on a detailed analysis of more than 2,000 detailed work activities (DWAs) tracked by the US Bureau of Labor Statistics (BLS). Against each of these activities, the analysis quantified the required level of performance in 18 individual capabilities.

For this report, we classified seven of these 18 capabilities as relevant to deep learning, in that these are capabilities deep learning is well suited to implement (Exhibit A20). For example, deep learning networks have dramatically improved the ability of machines to recognize images, which is a form of sensory perception. Thus, we include this in our list of seven deep learning capabilities: natural language understanding, sensory perception, generating novel patterns/categories, social and emotional sensing, recognizing known patterns/categories, optimization and planning, and natural language generation.

Exhibit A20

Deep learning is well suited to develop seven out of 18 capabilities required in many work activities



NOTE: While this example illustrates the activities performed by a retail worker only, we analyzed some 2,000 activities across all occupations.

SOURCE: McKinsey Global Institute analysis

We are able to quantify the potential wage impact of deep learning by linking the DWAs to the occupations that require them. For each occupation, we use the fraction of time spent on an activity to quantify the value associated with that DWA occupation pair. Then we aggregated across occupations to the DWA level to quantify the total value associated with that DWA.

For some exhibits, we use DWA groups to summarize the impact across the approximately 2,000 DWAs. These groups are provided by the BLS and classify the DWAs into 37 "elements" or categories. Since each DWA falls into one of these categories, any analysis can be aggregated up to the DWA group level by summing across impact or wages for the relevant DWAs.

BIBLIOGRAPHY

A

Andel, C., S. L. Davidow, M. Hollander, and D. A. Moreno, "The economics of health care quality and medical errors," *Journal of Health Care Finance*, volume 39, number 1, fall 2012.

B

Booth, Adrian, Niko Mohr, and Peter Peters, "The digital utility: New opportunities and challenges," McKinsey.com, May 2016.

Bughin, Jacques, "Big data, big bang?" *Journal of Big Data*, volume 3, number 2, January 2016.

Bughin, Jacques, "Big data: Getting a better read on performance," *The McKinsey Quarterly*, February 2016.

Bughin, Jacques, "Ten lessons learned from big data analytics," *Journal of Applied Marketing Analytics*, forthcoming.

C

Charles, Dustin, Meghan Gabriel, and Talisha Searcy, "Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2014," Office of the National Coordinator for Health Information Technology, data brief number 23, April 2015.

Chui, Michael, and James Manyika, "Competition at the digital edge: 'Hyperscale businesses,'" *The McKinsey Quarterly*, March 2015.

Chui, Michael, James Manyika, and Mehdi Miremadi, "Four fundamentals of workplace automation," *The McKinsey Quarterly*, November 2015.

D

Dills, Angela K., and Sean E. Mulholland, *Ride-sharing, fatal crashes, and crime*, Providence College working paper, May 2016.

Dilmegani, Cem, Bengi Korkmaz, and Martin Lundqvist, "Public-sector digitization: The trillion-dollar challenge," McKinsey.com, December 2014.

Discovery Health, *Sustainable development report*, 2015.

E

Ericsson, *Ericsson mobility report: On the pulse of the networked society*, June 2016.

G

Gantz, John, and David Reinsel, *The digital universe in 2020*, IDC, February 2013.

Gore, Adrian, "How Discovery keeps innovating," *The McKinsey Quarterly*, May 2015.

Greenwood, Brad N., and Sunil Wattal, *Show me the way to go home: An empirical investigation of ride sharing and alcohol related motor vehicle homicide*, Temple University, Fox School of Business research paper number 15-054, January 2015.

H

Hilbert, Martin, and Priscila López, "The world's technology capacity to store, communicate, and compute information," *Science*, volume 332, issue 6025, April 2011.

I

Indeed.com, *Beyond the talent shortage: How tech candidates search for jobs*, September 2015.

K

Kohn, Linda T., Janet M. Corrigan, and Molla S. Donaldson, eds., *To err is human: Building a safer health system*, Committee on Quality of Health Care in America, Institute of Medicine, National Academy Press, 2000.

Kuyer, Lior, Shimon Whiteson, Bram Bakker, and Nikos Vlassis, "Multiagent reinforcement learning for urban traffic control using coordination graphs," *Machine learning and knowledge discovery in databases*, volume 5211 of Lecture Notes in Computer Science series, Springer Berlin Heidelberg, 2008.

L

Lund, Susan, James Manyika, and Kelsey Robinson, "Managing talent in a digital age," *The McKinsey Quarterly*, March 2016.

M

Makary, Martin A., and Michael Daniel, "Medical error: The third leading cause of death in the US," *BMJ*, May 2016.

McKinsey Advanced Industries Practice, *Automotive revolution—perspective towards 2030*, January 2016.

McKinsey Advanced Industries Practice, *Car data: Paving the way to value-creating mobility*, March 2016.

McKinsey & Company survey, "The need to lead in data and analytics," McKinsey.com, April 2016, available at http://www.mckinsey.com/business-functions/business-technology/our-insights/the-need-to-lead-in-data-and-analytics.

McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity*, June 2011.

McKinsey Global Institute, *China's digital transformation: The internet's impact on productivity and growth*, July 2014.

McKinsey Global Institute, *Digital America: A tale of the haves and have-mores*, December 2015.

McKinsey Global Institute, *Digital Europe: Pushing the frontier, capturing the benefits*, June 2016.

McKinsey Global Institute, *Disruptive technologies: Advances that will transform life, business, and the global economy*, May 2013.

McKinsey Global Institute, *The internet of things: Mapping the value beyond the hype*, June 2015.

McKinsey Global Institute, *Open data: Unlocking innovation and performance with liquid information*, October 2013.

McKinsey Global Institute, *Playing to win: The new global competition for corporate profits*, September 2015.

Muehlhauser, Luke, and Lila Rieber, “Exponential and non-exponential trends in information technology,” Machine Intelligence Research Institute blog, May 12, 2014.

P

Pande, Vijay, “FAH’s achievements in 2015, with a glimpse into 2016,” Folding@home blog, December 6, 2015.

Patel, Tejal A., Mamta Puppala, Richard O. Ogunti, Joe Edward Ensor, Tiangchen He, and Jitesh Shewale, “Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods: Natural Language Processing,” *Cancer*, August 2016.

R

Ransbotham, Sam, David Kiron, and Pamela Kirk Prentice, “The talent dividend: Analytics talent is driving competitive advantage at data-oriented companies,” *MIT Sloan Management Review*, April 25, 2015.

Robb, Drew, “Are we running out of data storage space?” World Economic Forum blog, November 10, 2015.

S

SAP, *Real-time enterprise stories*, case studies from Bloomberg Businessweek Research Services and Forbes Insights, October 2014.

Stawicki, Elizabeth, “Talking scales and telemedicine: ACO tools to keep patients out of the hospital,” *Kaiser Health News*, August 2013.

U

US Department of Energy, *2014 smart grid system report*, August 2014.

RELATED MGI AND MCKINSEY RESEARCH

Big Data: The next frontier for innovation, competition, and productivity (May 2011)

Big data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus—as long as the right policies and enablers are in place.

The Internet of Things: Mapping the value beyond the hype (June 2015)

If policymakers and businesses get it right, linking the physical and digital worlds could generate up to $1.1 trillion a year in economic value by 2025.

Disruptive technologies: Advances that will transform life, business, and the global economy (May 2013)

Twelve emerging technologies—including the mobile Internet, autonomous vehicles, and advanced genomics—have the potential to truly reshape the world in which we live and work. Leaders in both government and business must not only know what's on the horizon but also start preparing for its impact.

Digital America: A tale of the haves and have-mores (December 2015)

While the most advanced sectors, companies, and individuals push the boundaries of technology use, the US economy as a whole is realizing only 18 percent of its digital potential.

www.mckinsey.com/mgi

E-book versions of selected MGI reports are available at MGI's website, Amazon's Kindle bookstore, and Apple's iBooks Store.

Download and listen to MGI podcasts on iTunes or at www.mckinsey.com/mgi/publications/multimedia/

Cover image: © Nadla/Getty Images.

Cover insets (left to right): Businesswomen in conference room © Hero Images/Getty Images, young man in city © svetikd/Getty Images, builders with tablet © Monty Rakusen/Getty Images.

Contents page images (top to bottom): Woman and man at computer © Hero Images/Getty Images, doctor with screen © Wichy/Shutterstock, man with face recognition scanner © Monty Rakusen/Getty Images.

McKinsey Global Institute
December 2016
Copyright © McKinsey & Company
www.mckinsey.com/mgi

@McKinsey_MGI
McKinseyGlobalInstitute